

# COGNITIVE



Edited by Nick Braisby and Angus Gellatly

# PSYCHOLOGY

Cognitive processes enable us to experience the world around us: to recognize familiar faces; to communicate to one another through speech and writing; to feel emotion as we recall memories from the past. *Cognitive Psychology* provides a dynamic and exciting insight into this illuminating subject, leading us through such topics as attention, memory, judgement and decision making, and introducing us to the latest computational and imaging techniques through which our understanding of these topics is being continually enhanced. With state-of-the-art introductions from internationally recognized experts, *Cognitive Psychology* conveys a sense of excitement in the subject which is sure to engage and enthuse any student.

- A unique emphasis on the role of methodologies, including imaging techniques, gives the student a real flavour of the excitement of contemporary cognitive psychology.
- The study of cognition is presented in the broader context of themes, challenges, and issues, not just theory, to stimulate deeper thought, and encourage the student to take a more critical approach to their subject.
- Rich pedagogy includes
  - links between the theory of cognitive psychology and aspects of everyday life, presenting material in contexts which students can relate to and understand
  - boxes that link research findings with the techniques used to elucidate them, to develop the student's awareness and appreciation of scientific method.
- Chapters from expert researchers active in their fields have been carefully edited to lead students through the subject in a coherent way.



[www.oup.com/uk/booksites/psychology](http://www.oup.com/uk/booksites/psychology)

The companion web site includes

- figures from the book available to download to facilitate lecture preparation
- experiments and data sets to enable students to explore material covered in an interactive, problem-based fashion.

*Cognitive Psychology: A Methods Companion* accompanies this text. It explores in more detail some of the key methods used or encountered by cognitive psychologists, including connectionism, neuroimaging, neuropsychology, statistical methods, and ethics.



The Open  
University

DD303 BOOK

OXFORD  
UNIVERSITY PRESS



# **Cognitive Psychology**



# **Cognitive Psychology**

Edited by Nick Braisby and Angus Gellatly

**OXFORD**  
UNIVERSITY PRESS

Oxford University Press in association with The Open University



Great Clarendon Street, Oxford OX2 6DP

Published by Oxford University Press, Oxford in association with The Open University,  
Milton Keynes



Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York Auckland Bangkok Buenos Aires Cape Town Chennai  
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata Kuala Lumpur Madrid  
Melbourne Mexico City Nairobi São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

Published in the United States by Oxford University Press Inc., New York

The Open University, Walton Hall, Milton Keynes, MK7 6AA

First published 2005.

Copyright © 2005 The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilized in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd of 90 Tottenham Court Road, London W1T 4LP.

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without similar condition including this condition being imposed on the subsequent purchaser.

British Library Cataloguing in Publication Data available.

Library of Congress Cataloguing in Publication Data available.

Edited, designed and typeset by The Open University.

Printed in the United Kingdom by Scotprint, Haddington.

ISBN 0-19-927376-6

# Preface

This book has been produced as the core text for the Open University's level 3 course in *Cognitive Psychology* (DD303). However, it has been designed to serve students taking other courses in cognitive psychology as well, either as essential or recommended reading. There are a number of features of the design of this text that we hope will serve well both students learning about cognitive psychology and educators teaching the subject.

## Book structure

The chapters in this book are organized in five parts. The first four parts focus on broad and well-established topic areas within cognitive psychology, such as perceptual processes and memory. The fifth considers a range of challenges, themes and issues – topics that have been thought to present challenges to the cognitive approach, such as emotion and consciousness; themes such as cognitive modelling and modularity; and issues such as the relation of cognition to biology.

The first chapter is not located in one of these parts. It attempts to give a historical and conceptual introduction to cognitive psychology, laying out the foundations of the subject, and raising some of the important themes and issues that are revisited in later chapters. Some of these themes are developed also in the introductions to each of the subsequent parts; we recommend that students read these introductions prior to reading their associated parts, and re-read them afterwards.

## Chapter structure

Each chapter has been structured according to certain conventions.

An **emboldened term** signifies the introduction of a key concept or term that is either explicitly or implicitly defined in the surrounding text. The locations of these defined terms are also flagged in bold in the index.

Each chapter contains a number of **activities**. Often these may be simple thought exercises that may take no more than a minute or so. Others are more involved. Each activity has been integrated into the design of the chapter, and is aimed at enhancing students' understanding of the material. We recommend that student readers attempt as many of these activities as possible and, where appropriate, revisit them after completing each chapter.

The chapters in this book also make use of **text boxes**. Each box has been written to amplify a particular aspect of the material without interrupting the ongoing narrative. Though the boxes illuminate a wide range of issues, many focus on aspects of **research studies** and **methods**. Students may find they wish to finish a section before reading a particular box.

Each substantive main section finishes with a **section summary**, often a bullet point list reminding the student of the key points established in that section. We hope that students will use these as useful barometers of their understanding and re-read sections where the summary points are not clearly understood.

Each chapter makes a number of explicit **links** to other chapters in the book, often to specific numbered sections. It would be tedious in the extreme to continually follow each and every link, flicking to the relevant pages and reading the relevant 'linked' section. Rather, these links are intended to help students perceive the interconnected nature of cognitive psychology, identifying connections between topics that otherwise

might seem disparate. Of course, we hope that students will be motivated to follow some of these links either on first reading, or on a later reading, perhaps as a revision aid.

As well as a list of references, each chapter ends with some specific suggestions for **further reading**. While each chapter is designed to be self-contained, inevitably some issues get less attention than they deserve, and so interested readers may wish to pursue some of these suggestions for a more in-depth treatment. Moreover, it is always worth approaching a topic from more than one direction – consulting different texts, including other general texts on cognitive psychology, can help achieve a richer understanding and we recommend this approach to all students.

### **Supporting a course in cognitive psychology**

There are few restrictions on how one might use this text to support the teaching of a course in cognitive psychology. The chapters in this book may be tackled in a number of different orders. Depending on the focus of the course, particular parts may be omitted, or particular chapters omitted from a given part or parts. The book as a whole presupposes relatively little prior knowledge of cognitive psychology on the part of a student. However, in some instances, later chapters may presuppose some limited knowledge of related earlier chapters, though this is usually explicitly indicated. Similarly, while all chapters are designed to be taught at the same level, later chapters may tackle issues considered too complex in the earlier chapters. By focusing more on earlier or later chapters, courses can vary somewhat the degree of difficulty of the material they present.

### **Companion volume**

Accompanying this book is a companion publication *Cognitive Psychology: A Methods Companion*, also published by Oxford University Press and also designed as a key teaching text for the Open University's level 3 course in *Cognitive Psychology*. The *Methods Companion* considers in detail a number of key methodological issues in cognitive psychology, including ethics, connectionism, symbolic modelling, neuroimaging, neuropsychology and statistics.

### **Companion web site**

This book and the *Methods Companion* are associated with a companion web site that contains much additional material that can be used to further students' understanding and may be used in presenting a course in cognitive psychology ([www.oup.com/uk/booksites/psychology](http://www.oup.com/uk/booksites/psychology)). Materials include electronic versions of figures, experiment and data files, and software for running cognitive models.

### **Acknowledgements**

Finally, developing the Open University's level 3 course in *Cognitive Psychology* (DD303) has been a major undertaking, involving the production of two books, various pieces of software and associated files, audio materials, web sites and web-based materials, and numerous other additional items and activities. To say that such a course, and that this text, could not have been produced without the help and cooperation of a large number of people is an understatement. The following page lists those who have made this enterprise possible, and to each we extend our grateful and sincere thanks, as we do to anyone we have omitted in error.

Nick Braisby and Angus Gellatly, Milton Keynes, January 2005

# Cognitive Psychology Course Team

This book was designed and produced for The Open University course DD303 *Cognitive Psychology*. The editors gratefully thank all those people, listed below, who have been involved in the process (based at The Open University, unless otherwise stated).

## CORE COURSE TEAM:

**Course Chair:** Nick Braisby

**Course Manager:** Ingrid Slack

**Core Team Members:** Sandy Aitkenhead; Nicola Brace; Angus Gellatly; Alison J.K. Green; Martin Le Voi; Bundy Mackintosh; Peter Naish; Graham Pike

**Course Manager (rights):** Ann Tolley

**Course Secretaries:** Marie Morris; Elaine Richardson

**Additional Authors:** Jackie Andrade (University of Sheffield); Peter Ayton (City University; Chris Barry (University of Essex); Simon Bignell (University of Essex); Martin A. Conway (University of Durham); Graham Edgar (University of Gloucestershire); Simon Garrod (University of Glasgow); Gareth Gaskell (University of York); Ken Gilhooly (University of Paisley); Olaf Hauk (MRC Cognition and Brain Sciences Unit); Graham J. Hitch (University of York); Emily A. Holmes (MRC Cognition and Brain Sciences Unit); Ashok Jansari (University of East London); Helen Kaye; Paul Mulholland; Mike Oaksford (Cardiff University); Mike Pilling; John Richardson; Andrew Rutherford (Keele University); Anthony J. Sanford (University of Glasgow); Stella Tickle; Tony Stone (London South Bank University); Stuart Watt (Robert Gordon University); Jenny Yiend (MRC Cognition and Brain Sciences Unit)

**Course Reader:** Matt Lambon Ralph (University of Manchester)

**External Assessor:** James Hampton (City University)

**Media Project Manager:** Lynne Downey

**Production and Presentation Administrator:** Richard Golden

**Copublishing Adviser:** Jonathan Hunt

**Lead Editor:** Chris Wooldridge

**Editors:** Alison Edwards; Kathleen Calder; Winifred Power (Freelance)

**Designers:** Tammy Alexander; Alison Goslin; Diane Mole

**Graphic Artists:** Janis Gilbert; Sara Hack

**Picture Researcher:** Celia Hart

**eMedia Quality Promoter:** Roger Moore

**Software Designers:** Ian Every; Maurice Brown; David Morris

**Rights Adviser:** Alma Hales

**Contracts Executives:** Katie Meade; Sarah Gamman

**Compositors:** Pam Berry; Lisa Hale; Phillip Howe

**Print Buyer Controller:** Lene Connolly

**Assistant Print Buyer:** Dave Richings

This publication forms part of an Open University course DD303 *Cognitive Psychology*. Details of this and other Open University courses can be obtained from the Course Information and Advice Centre, PO Box 724, The Open University, Milton Keynes MK7 6ZS, United Kingdom: tel. +44 (0)1908 653231, e-mail general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at <http://www.open.ac.uk> where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials visit the webshop at [www.ouw.co.uk](http://www.ouw.co.uk), or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure. tel. +44 (0)1908 858785; fax +44 (0)1908 858787; e-mail ouwenq@open.ac.uk

# Contents in brief

<b>1: Foundations of cognitive psychology</b> <i>Nick Braisby and Angus Gellatly</i>	1
<hr/>	
<b>PART 1 PERCEPTUAL PROCESSES</b>	
Introduction	34
<b>2: Attention</b> <i>Peter Naish</i>	37
<b>3: Perception</b> <i>Graham Pike and Graham Edgar</i>	71
<b>4: Recognition</b> <i>Graham Pike and Nicola Brace</i>	113
<hr/>	
<b>PART 2 CONCEPTS AND LANGUAGE</b>	
Introduction	158
<b>5: Concepts</b> <i>Nick Braisby</i>	163
<b>6: Language processing</b> <i>Gareth Gaskell</i>	197
<b>7: Language in action</b> <i>Simon Garrod and Anthony J. Sanford</i>	231
<hr/>	
<b>PART 3 MEMORY</b>	
Introduction	266
<b>8: Long-term memory: encoding to retrieval</b> <i>Andrew Rutherford</i>	269
<b>9: Working memory</b> <i>Graham J. Hitch</i>	307
<hr/>	
<b>PART 4 THINKING</b>	
Introduction	344
<b>10: Problem solving</b> <i>Alison J.K. Green and Ken Gilhooly</i>	347
<b>11: Judgement and decision making</b> <i>Peter Ayton</i>	382
<b>12: Reasoning</b> <i>Mike Oaksford</i>	418
<hr/>	
<b>PART 5 CHALLENGES, THEMES AND ISSUES</b>	
Introduction	458
<b>13: Cognition and emotion</b> <i>Jenny Yiend and Bundy Mackintosh</i>	463
<b>14: Autobiographical memory and the working self</b> <i>Martin A. Conway and Emily A. Holmes</i>	507
<b>15: Consciousness</b> <i>Jackie Andrade</i>	545
<b>16: Cognitive modelling and cognitive architectures</b> <i>Paul Mulholland and Stuart Watt</i>	579
<b>17: Theoretical issues in cognitive psychology</b> <i>Tony Stone</i>	617
Epilogue	655
Index	659
Acknowledgements	684



# Contents

<b>Chapter 1: Foundations of cognitive psychology</b>	
<i>Nick Braisby and Angus Gellatly</i>	
<b>1 Introduction</b>	1
<b>2 What is cognitive psychology?</b>	2
<b>3 A brief history of cognitive psychology</b>	8
3.1 Introspectionism	8
3.2 Gestalt psychology	9
3.3 Behaviourism	10
3.4 The return of the cognitive	12
<b>4 Science, models and the mind</b>	19
<b>5 The cognitive approach</b>	22
5.1 Representation	22
5.2 Computation	24
<b>6 Level-dependent explanations</b>	27
6.1 The computational level	27
6.2 The algorithmic level	27
6.3 The implementational level	29
6.4 Using Marr's levels	29
<b>7 Conclusions</b>	30
<b>Further reading</b>	31
<b>References</b>	31
<hr/> <b>PART 1 PERCEPTUAL PROCESSES</b>	
Introduction	34
<b>Chapter 2: Attention</b> <i>Peter Naish</i>	
<b>1 Auditory attention</b>	37
1.1 Disentangling sounds	37
1.2 Attending to sounds	41
1.3 Eavesdropping on the unattended message	43
<b>2 Visual attention</b>	45
2.1 Knowing about unseen information	46
2.2 Towards a theory of parallel processing	49
2.3 Rapid serial visual presentation	50
2.4 Masking and attention	54
<b>3 Integrating information in clearly-seen displays</b>	55
3.1 Serial and parallel search	55
3.2 Non-target effects	56
3.3 The 'flanker' effect	57

<b>4</b>	<b>Attention and distraction</b>	59
4.1	The effects of irrelevant speech	60
4.2	Attending across modalities	61
<b>5</b>	<b>The neurology of attention</b>	62
5.1	The effects of brain damage	62
5.2	Event-related potentials	64
<b>6</b>	<b>Concluding thoughts</b>	65
	<b>Further reading</b>	67
	<b>References</b>	67
 <b>Chapter 3: Perception</b> <i>Graham Pike and Graham Edgar</i>		
<b>1</b>	<b>Introduction</b>	71
1.1	Perceiving and sensing	73
1.2	The eye	74
1.3	Approaches to perception	75
<b>2</b>	<b>The Gestalt approach to perception</b>	77
<b>3</b>	<b>Gibson's theory of perception</b>	80
3.1	An ecological approach	81
3.2	The optic array and invariant information	82
3.3	Flow in the ambient optic array	86
3.4	Affordances and resonance	89
<b>4</b>	<b>Marr's theory of perception</b>	90
4.1	The grey level description	91
4.2	The primal sketch	92
4.3	The 2½D sketch	95
4.4	Evaluating Marr's approach	96
<b>5</b>	<b>Constructivist approaches to perception</b>	98
<b>6</b>	<b>The physiology of the human visual system</b>	102
6.1	From the eye to brain	102
6.2	The dorsal and ventral streams	103
6.3	The relationship between visual pathways and theories of perception	104
6.4	A dual-process approach?	105
6.5	Combining bottom-up and top-down processing	106
<b>7</b>	<b>Conclusion</b>	108
	<b>Further reading</b>	109
	<b>References</b>	109
 <b>Chapter 4: Recognition</b> <i>Graham Pike and Nicola Brace</i>		
<b>1</b>	<b>Introduction</b>	113
1.1	Recognition in the wider context of cognition	114

<b>2</b>	<b>Different types of recognition</b>	115
2.1	Object and face recognition	115
2.2	Active processing – recognizing objects by touch	118
2.2	Recognizing two-dimensional objects	120
2.3	Object-centred vs viewer-centred descriptions	122
<b>3</b>	<b>Recognizing three-dimensional objects</b>	124
3.1	Marr and Nishihara's theory	124
3.2	Evaluating Marr and Nishihara's theory	131
3.3	Biederman's theory	131
<b>4</b>	<b>Face recognition</b>	135
4.1	Recognizing familiar and unfamiliar faces	136
<b>5</b>	<b>Modelling in face recognition</b>	138
5.1	A connectionist model of face recognition	141
<b>6</b>	<b>Neuropsychological evidence</b>	144
<b>7</b>	<b>Are faces 'special'?</b>	148
<b>8</b>	<b>Conclusion</b>	153
	<b>Further reading</b>	153
	<b>References</b>	153
<b>PART 2</b>	<b>CONCEPTS AND LANGUAGE</b>	
	Introduction	158
	<b>Chapter 5: Concepts</b> <i>Nick Braisby</i>	
<b>1</b>	<b>Introduction</b>	163
1.1	Concepts, categories and words	163
1.2	Categorization	164
1.3	The wider story of concepts	166
1.4	Concepts and cognition	167
<b>2</b>	<b>Explaining categorization</b>	169
2.1	Similarity I: the classical view of concepts	169
2.2	Similarity II: prototype theories of concepts	175
2.3	Common-sense theories: the theory-based view	180
2.4	Psychological essentialism	184
<b>3</b>	<b>Where next?</b>	187
3.1	Is all categorization the same?	188
3.2	Are all concepts the same?	189
3.3	Are all categorizers the same?	190
<b>4</b>	<b>Conclusion</b>	191
	<b>Further reading</b>	192
	<b>References</b>	192

<b>Chapter 6: Language processing</b>	<i>Gareth Gaskell</i>	
<b>1</b>	<b>Introduction</b>	197
<b>2</b>	<b>Word recognition</b>	198
2.1	Spoken word recognition	198
2.2	Visual word recognition	206
<b>3</b>	<b>The mental lexicon</b>	213
3.1	Morphology	213
3.2	Accessing word meanings	215
<b>4</b>	<b>Sentence comprehension</b>	219
4.1	Syntax	220
4.2	Models of parsing	222
4.3	Is parsing autonomous?	224
4.4	Constraints on parsing	225
<b>5</b>	<b>Conclusion</b>	226
	<b>Further reading</b>	227
	<b>References</b>	228
<b>Chapter 7: Language in action</b>	<i>Simon Garrod and Anthony J. Sanford</i>	
<b>1</b>	<b>Introduction</b>	231
<b>2</b>	<b>Written language and discourse</b>	232
2.1	Processes underlying text interpretation	233
2.2	Special topics in understanding text	240
<b>3</b>	<b>Language production as a self-contained process</b>	245
3.1	Speech errors and the architecture of the language production system	245
3.2	Message selection and audience design	249
3.3	Self-monitoring	251
<b>4</b>	<b>The challenge of dialogue</b>	253
4.1	What is dialogue?	253
4.2	Dialogue and consensus	255
4.3	A model of dialogue processing	256
<b>5</b>	<b>The monologue /dialogue distinction and group decision making</b>	258
<b>6</b>	<b>Summary</b>	260
	<b>Further reading</b>	261
	<b>References</b>	261
<b>PART 3</b>	<b>MEMORY</b>	
	Introduction	266
<b>Chapter 8: Long-term memory: encoding to retrieval</b>	<i>Andrew Rutherford</i>	

<b>1</b>	<b>Introduction</b>	269
<b>2</b>	<b>Encoding</b>	270
2.1	Levels of processing	270
2.2	Relational and item-specific processing	273
<b>3</b>	<b>Memory stores and systems</b>	277
3.1	Multiple memory systems	278
3.2	Declarative and procedural memory	282
<b>4</b>	<b>Retrieval</b>	284
4.1	Encoding specificity and transfer appropriate processing	284
<b>5</b>	<b>Implicit memory</b>	286
5.1	Perceptual and conceptual memory	286
5.2	Accounts of implicit memory	288
5.3	Implicit memory and amnesia	290
<b>6</b>	<b>Jacoby's process-dissociation framework</b>	292
<b>7</b>	<b>Remember and know judgements</b>	295
7.1	Do remember and know judgements reflect different response criteria?	297
<b>8</b>	<b>Conclusions</b>	299
	<b>Further reading</b>	300
	<b>References</b>	300

**Chapter 9: Working memory** *Graham J. Hitch*

<b>1</b>	<b>Introduction</b>	307
1.1	Human memory as a multifaceted system	307
1.2	Distinction between short-term and long-term memory	308
1.3	Working memory as more than STM	310
<b>2</b>	<b>The structure of working memory</b>	314
2.1	A multi-component model	314
2.2	Phonological working memory	317
2.3	Executive processes	323
<b>3</b>	<b>Vocabulary acquisition</b>	329
3.1	Neuropsychological evidence	329
3.2	Individual differences	330
3.3	Experimental studies	330
<b>4</b>	<b>Modelling the phonological loop</b>	331
4.1	Serial order	332
<b>5</b>	<b>Conclusion</b>	335
	<b>Further reading</b>	336
	<b>References</b>	336

<b>PART 4 THINKING</b>	
Introduction	344
<b>Chapter 10: Problem solving Alison J.K. Green and Ken Gilhooly</b>	
1 Introduction	347
1.1 What is a 'problem'	349
1.2 Protocol analysis in problem-solving research	350
2 'Simple' problem solving	353
2.1 The Gestalt legacy	353
2.2 Representation in puzzle problem solving	356
2.3 The information processing approach: problem solving as search	358
2.4 Information processing approaches to insight	361
3 Analogical problem solving	363
3.1 Analogies in problem solving	363
3.2 How do analogies work?	364
4 'Complex' problem solving	365
4.1 The role of knowledge in expert problem solving	366
4.2 A modal model of expertise?	370
5 Prospects for problem-solving research	371
5.1 Does expertise transfer?	371
5.2 Individual differences	372
6 Conclusion	376
Further reading	377
References	377
<b>Chapter 11: Judgement and decision making Peter Ayton</b>	
1 Introduction	382
1.1 Theories of decision making	383
1.2 Supporting decision making	383
2 Normative theory of choice under risk	384
2.1 Prescriptive application of normative theory: decision analysis	385
2.2 Axioms underlying subjective expected utility theory	387
2.3 Violations of the axioms	388
3 Findings from behavioural decision research	392
3.1 The 'preference reversal phenomenon'	393
3.2 Causes of anomalies in choice	394

<b>4</b>	<b>Prospect theory</b>	396
4.1	Prospect theory and 'loss aversion'	398
4.2	'Framing' effects	399
<b>5</b>	<b>Judgement under uncertainty</b>	400
5.1	Judging probabilities and Bayes' Theorem	400
5.2	Does Bayes' Theorem describe human judgement?	402
5.3	Heuristics and biases	404
5.4	Evaluating the heuristics and biases account	406
5.5	Overconfidence	408
<b>6</b>	<b>Fast and frugal theories of decision making</b>	410
<b>7</b>	<b>Conclusion</b>	412
	<b>Further reading</b>	413
	<b>References</b>	413
 <b>Chapter 12: Reasoning</b> <i>Mike Oaksford</i>		
<b>1</b>	<b>Introduction</b>	418
1.1	Reasoning and logic	418
1.2	Reasoning in everyday life	419
<b>2</b>	<b>Deductive reasoning and logic</b>	421
2.1	Logical connectives	421
2.2	When are arguments logically valid?	421
2.3	Logically invalid inferences	423
2.4	Form and meaning in logic	424
<b>3</b>	<b>Psychological theories of reasoning</b>	425
3.1	Mental logic	425
3.2	Mental models	425
3.3	The probabilistic approach	425
<b>4</b>	<b>Conditional inference</b>	427
4.1	The abstract conditional inference task	427
4.2	Everyday reasoning and the suppression effect	432
<b>5</b>	<b>Wason's selection task</b>	437
5.1	The abstract selection task	437
5.2	The deontic selection task	443
<b>6</b>	<b>Conclusion</b>	448
6.1	Theoretical evaluation	449
6.2	Integration, dual processes and individual differences	451
	<b>Further reading</b>	452
	<b>References</b>	452

<b>PART 5 CHALLENGES, THEMES AND ISSUES</b>	
Introduction	458
<b>Chapter 13: Cognition and emotion Jenny Yiend and Bundy Mackintosh</b>	
1 Introduction	463
1.1 Components of emotion	464
2 Different emotions	469
2.1 Basic emotions	469
2.2 Verbal labels	473
2.3 The dimensional approach	474
3 The function of emotions	476
3.1 Emotions alter goals	477
3.2 Emotions mobilize physiological resources	478
3.3 Emotional expressions as communication	479
3.4 Emotions as information	479
3.5 What is the function of emotional feelings?	481
4 Emotion influences cognition	481
4.1 Some important concepts	481
4.2 Memory	483
4.3 Attention	488
4.4 Semantic interpretation	491
5 Does cognition influence emotion?	494
5.1 A look at some historical answers	494
5.2 A clash of minds: the cognition/emotion debate	499
6 General summary	502
Further reading	503
References	503
<b>Chapter 14: Autobiographical memory and the working self Martin A. Conway and Emily A. Holmes</b>	
1 What are autobiographical memories?	507
2 Autobiographical memory across the lifespan	509
2.1 Childhood amnesia	511
2.2 The reminiscence bump	512
2.3 Recency	513
3 Autobiographical knowledge, episodic memory, the working self and memory construction	514
3.1 Autobiographical knowledge	517
3.2 Episodic and semantic memory	520
3.3 The working self	522
3.4 Constructing autobiographical memories	525

<b>4</b>	<b>Autobiographical memory in distress</b>	529
4.1	Traumatic event	531
4.2	Response at the time of trauma	531
4.3	Subsequent psychological symptoms	532
4.4	Impact of symptoms	534
4.5	The nature of intrusive trauma memories	535
<b>5</b>	<b>Conclusion: what are autobiographical memories for?</b>	537
	<b>Further reading</b>	538
	<b>References</b>	538

**Chapter 15: Consciousness** Jackie Andrade

<b>1</b>	<b>Introduction</b>	545
1.1	Defining consciousness	546
1.2	Philosophical approaches to consciousness	548
1.3	The place of consciousness within cognitive psychology	550
<b>2</b>	<b>Empirical research: cognitive studies of consciousness</b>	552
2.1	Implicit cognition	552
2.2	Controlled versus automatic processing	560
2.3	The neuropsychology of consciousness	562
<b>3</b>	<b>What is consciousness for?</b>	564
3.1	Consciousness and behavioural control	564
3.2	Cross-talk between cognitive modules	568
3.3	Altered states of consciousness	569
<b>4</b>	<b>Cognitive theories of consciousness</b>	571
<b>5</b>	<b>Conclusion: what can cognitive psychology tell us about consciousness?</b>	574
	<b>Further reading</b>	574
	<b>References</b>	575

**Chapter 16: Cognitive modelling and cognitive architectures** Paul Mulholland and Stuart Watt

<b>1</b>	<b>What is cognitive modelling?</b>	579
1.1	Parallel distributed processing	579
1.2	Rule based systems	583
1.3	Cognitive architectures	584
<b>2</b>	<b>An overview of ACT-R</b>	585
2.1	A brief history of ACT-R	585
2.2	The architecture of ACT-R	586
2.3	Declarative memory	587
2.4	Procedural memory	588

2.5	Goals and the goal stack	590
<b>3</b>	<b>ACT-R accounts of memory phenomena</b>	592
3.1	Declarative representation of lists	593
3.2	Production rules for the rehearsal and retrieval of lists	596
3.3	List activation	597
3.4	Running the model	599
3.5	Evaluation of the ACT-R approach to modelling memory	600
<b>4</b>	<b>Learning and using arithmetic skills</b>	601
4.1	Production compilation	601
4.2	An example of human problem-solving behaviour: addition by counting	604
4.3	Models of learning and problem solving in practice	607
<b>5</b>	<b>A comparison of ACT-R and PDP</b>	608
<b>6</b>	<b>When is a model a good model?</b>	611
<b>7</b>	<b>Conclusions</b>	614
	<b>Further reading</b>	614
	<b>References</b>	614
<b>Chapter 17: Theoretical issues in cognitive psychology</b>		
<i>Tony Stone</i>		
<b>1</b>	<b>Introduction</b>	617
<b>2</b>	<b>Computation and cognition</b>	619
2.1	Some basic ideas	619
2.2	Connectionism versus the CMM: the past-tense debate	621
<b>3</b>	<b>Modularity</b>	632
3.1	An outline of Fodor's theory of modularity	632
3.2	The central systems	637
3.3	Debates about modularity	639
<b>4</b>	<b>Cognitive psychology and the brain</b>	643
4.1	Levels of explanation	643
4.2	The co-evolution of cognitive and neurobiological theories	644
4.3	The radical neuron doctrine	646
<b>5</b>	<b>Conclusion</b>	649
	<b>Further reading</b>	650
	<b>References</b>	650
	Epilogue	655
	Index	659
	Acknowledgements	684

# Foundations of cognitive psychology

## Chapter 1

Nick Braisby and Angus Gellatly

---

### 1 Introduction

How does memory work? How do we understand language, and produce it so that others can understand? How do we perceive our environment? How do we infer from patterns of light or sound the presence of objects in our environment, and their properties? How do we reason, and solve problems? How do we think?

These are some of the foundational questions that cognitive psychology examines. They are foundational partly because each concerns the nature of a basic psychological ability, abilities that we often take for granted, yet which are vital to our normal, healthy functioning and are key to our understanding of what it means to be human. And they are foundational partly because they are important for psychology as a whole, and not just cognitive psychology. For instance, how can we hope to understand completely the behaviour of employees in an organization unless we first understand their perceptions and memories, and how they reason and attempt to solve problems? How can we understand the way in which people interact to shape one another's opinions if we do not understand how people understand and process language, and how they make judgements?

Throughout this book, the various authors tackle these and other questions, and show you how much of these foundations cognitive psychologists have so far uncovered. The book begins with an exploration of perceptual processes, moves to a discussion of categorization and language, through to memory, and then to thinking processes. The last part of the book is devoted to wider issues: to topics that have been thought to present a challenge to cognitive psychology – such as consciousness and emotion – and to some of the themes and theoretical questions which pervade the cognitive approach.

In this chapter, we try to answer the question ‘What is cognitive psychology?’ and, in so doing, outline some of the foundational assumptions that cognitive psychologists tend to make, as well as some of the reasons why it is such an important and fascinating subject – not least the fact that it raises many deep and important questions concerning the mind. We consider some of the issues that have attracted and continue to attract the interest of cognitive psychologists, and some of the assumptions they make in order to develop models and theories. We also consider the cognitive approach in general and the kinds of explanation cognitive psychologists favour. We touch upon the relations between cognitive psychology and other sub-disciplines of psychology, and those between cognitive psychology and other disciplines (such as philosophy, computing, and linguistics).

There are many substantial issues that we only touch on – it is not easy to define the relationship between two academic disciplines, for example – and so we only hope to convey something of their flavour here. Our aim in this chapter is therefore merely to *introduce* cognitive psychology, to explain some of its key distinguishing features, and to uncover some of the many broad issues lying beneath its surface.

You will obtain a richer and more complete overview of cognitive psychology from reading subsequent chapters, and especially Chapter 17. You may find that the current chapter raises as many questions as it answers and that, as your reading of this book progresses, you periodically want to revisit this chapter to gain a better understanding of issues that, on first reading, seemed hazy. If this chapter were only to raise questions that you have in mind when you read subsequent chapters, and to arouse your curiosity sufficiently that you periodically revisit this chapter, it will have served its purpose well.

## 2 What is cognitive psychology?

What is cognitive psychology? Well, as with most questions, there can be short or long answers. The short, though not uncontentious, answer is that cognitive psychology is the branch of psychology devoted to the scientific study of the mind. Straightforward as this may seem, to understand the nature of cognitive psychology means digging deeper. And it is an excavation that raises all manner of substantial and interesting issues – as diverse as the nature of normality and computation, and the importance of individual differences and brain images.

### ACTIVITY 1.1

---

Given the above definition that cognitive psychology is the scientific study of the mind, take a few minutes to write down some of what you would expect its characteristic features to be. For example, you might want to list what you take to be the characteristic features of a ‘scientific’ approach within psychology generally; and you might want to list some of the characteristic topics you would expect cognitive psychologists to study.

Keep your list ready to refer to as you read the rest of this chapter.

---

Activity 1.1 raises a number of interesting questions about the nature and scope of cognitive psychology. What does it mean for a psychology to be ‘cognitive’, for example? Did your list make any reference to normality? Well, when we say that cognitive psychology is the scientific study of the mind, this usually means ‘normally functioning human minds’. We can develop an understanding of the normal human mind in various ways: by studying people with normal minds and normal brains, for example; but also by studying people with abnormal minds or abnormal brains too, by studying animals of other species, and even devices, such as computers, with no brain at all. With respect to just this one issue – normality – cognitive psychology is clearly a broad enterprise. Box 1.1 gives a brief illustration of how evidence from people with brain damage can inform our understanding of normal cognition. Don’t worry too much if you cannot follow all of the details at this stage – just try to get a feel for how cognitive psychologists have tried to relate evidence from brain-damaged patients to normal cognition.

**1.1****Research study****Category-specific impairments I: neuropsychological methods**

Warrington and Shallice (1984) describe four patients with specific impairments in recognizing living things. Because the impairment was thought to be specific to the category of living things, it has been called a **category-specific impairment**. One patient, JBR, for example, experienced brain damage after suffering from herpes simplex encephalitis. As a result, when asked to name pictures, he correctly named only approximately 6 per cent of the pictures of living things, yet around 90 per cent of the pictures of non-living things. Other patients, though fewer of them, have been found to show an opposite impairment – that is, an impairment primarily to the category of non-living things (Hillis and Caramazza, 1991).

These studies have suggested to researchers that, in normal cognition, the categories of living and non-living things might be represented and/or processed differently. For example, one suggestion, that has since been much debated, has been that in normal cognition the functional and sensory properties of categories are represented differently, and that living things tend to depend more on the sensory properties, while non-living things depend more on functional properties (Warrington and Shallice, 1984). The suggestion was also at first thought to help explain why JBR, on the assumption that he has an impairment for sensory properties, was also found to show impairments for some non-living categories, such as the categories of musical instruments and foods.

'Cognitive psychology' can also be used to refer to activities in a variety of other disciplines and sub-disciplines (did your list refer to other disciplines?). Some sub-disciplines, like cognitive neuropsychology, developmental cognitive neuropsychology, cognitive neuropsychiatry, and cognitive neuroscience, include the cognitive signifier in their own titles. Others, such as behavioural neurobiology, linguistics and artificial intelligence, do not; and some practitioners of these might well object to finding themselves included under the cognitive psychology umbrella. As you will see in Chapter 5, uncertainty and negotiation regarding membership are characteristic of many if not all of our conceptual categories. Our advice is not to worry too much about such definitional issues at this stage, and perhaps not even later on. But one thing that is clear is that there is no easily identified boundary between cognitive psychology and work carried on in other disciplines with which cognitive psychologists frequently engage.

Your list of features of cognitive psychology may have referred to some of the methods that cognitive psychologists employ: experiments, models (including computer models), neuropsychological investigations, and neuroimaging (or brain scans). Box 1.2 (overleaf) continues the discussion of category-specific impairments, and describes a study that combines features of experimental and neuroimaging methods.

**1.2****Research study****Category-specific impairments II: experimental and neuroimaging methods**

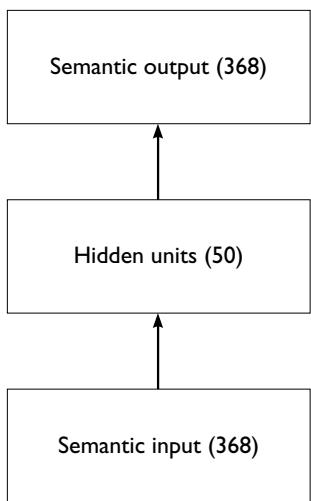
Devlin *et al.* (2000) combined features of experimental and neuroimaging methods to investigate whether the categories of living and non-living things could be associated with representations in different parts of the brain. One technique they used was a lexical decision task. In this task, participants either hear or see strings of letters (e.g. they might see the strings ‘warnd’ or ‘world’) and have to judge whether each string is a word or not. Experimenters typically record both the judgment made and the amount of time participants take to make their response (perhaps by pressing the appropriate button on a keyboard or response pad). Another task, that Devlin *et al.* called a semantic categorization task, required participants, having seen three words presented one after another, to judge whether a fourth word belonged to the same category as the first three. Devlin *et al.* carefully matched words for word frequency and letter length. Whilst performing the lexical decision and semantic categorization tasks described above, participants were scanned using positron emission tomography (PET) technology. Another group of participants performed the semantic categorization task using pictures that were matched for visual complexity; these participants were scanned using functional magnetic resonance imaging (fMRI) technology. Both of these scanning technologies enable experimenters to identify regions of the brain that are particularly active during the performance of a task. Critically, Devlin *et al.* found no differences between the categories of living and non-living things in terms of active regions of the brain in either the PET study or the fMRI study (see colour Plates 1 and 2). So the differences in representation discussed in Box 1.1 may not be associated with different brain regions (or perhaps these techniques were not sensitive enough to detect such differences).

Box 1.3 describes a study employing cognitive modelling methods to examine category-specific impairments.

**1.3****Research study****Category-specific impairments III: cognitive modelling**

Greer *et al.* (2001) developed a computational model based on the assumption that living things and non-living things were not represented in qualitatively distinct ways, but differences between them arise because living things have many shared properties that are strongly correlated (all mammals breathe, have eyes, etc.), whereas the properties of non-living things tend to be more distinctive. Greer *et al.* developed a form of computational model, called a connectionist network, which encoded these differences between living and non-living things. The model contained three kinds of units organized in three layers, as shown in Figure 1.1.

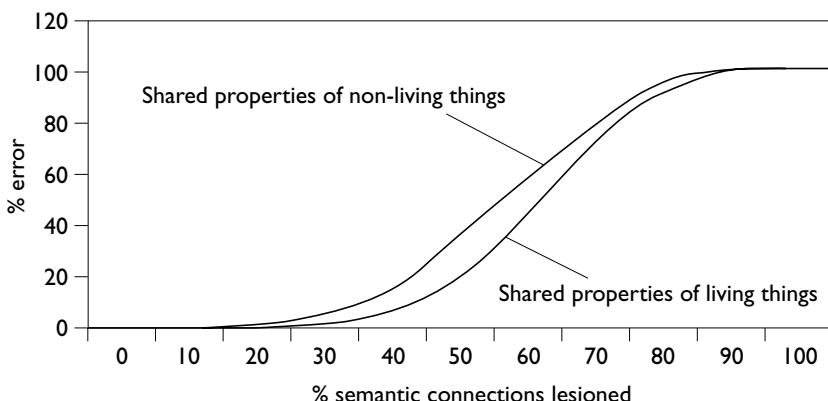




**Figure I.1** Architecture of Greer *et al.*'s connectionist network. The semantic input layer represents properties of categories. The network was trained until it could reproduce in the output layer the same pattern presented to its input layer. Arrows imply that every unit in a layer is connected to every unit in the subsequent layer. Numbers indicate the number of units in each layer

Source: Tyler and Moss, 2001, Figure I, p.248

However, information about the categories was distributed over the network's units in such a way that it was not possible to associate individual units with either living or non-living things. Greer *et al.* then artificially lesioned or damaged their network by removing 10 per cent of the network's connections at a time. They found that the shared properties of living things were more impervious to damage than those of non-living things, as shown in Figure I.2.



**Figure I.2** The results of 'lesioning' the model to simulate brain damage. As predicted by Greer *et al.*, the shared properties of living things were better preserved than the shared properties of non-living things, owing to the greater correlations between them

Source: Tyler and Moss, 2001, Figure II, p.249

Boxes 1.1 to 1.3 illustrate some of the methods that will be referred to throughout this book, and about some of which we will say more later. But, perhaps more obvious than any of these issues, Activity 1.1 raises the question of the subject matter of cognitive psychology. What is it that cognitive psychologists study?

An easy way of answering the question (and one you might have adopted for Activity 1.1) is scanning this book's table of contents. This will give you a good idea of the topics cognitive psychologists typically study, as, of course, will previous study of psychology. Certainly, the topics of perception, attention, language, categorization, reasoning, problem solving, and memory are central to the study of cognition. And cognition has broadened to include topics that have not always been seen as readily amenable to a cognitive approach (e.g. consciousness and emotion). The subsequent chapters will have much more to say about these issues than we can here. Activity 1.2 provides another way of thinking about the topics that interest cognitive psychologists.

## ACTIVITY 1.2

---

At this moment your behaviour involves getting information from this book. Your eyes may be scanning across the page and detecting patterns of colour, and light and shade; or, if you are listening to this book on audio CD or it is being machine-read from an electronic copy, your ears will be detecting sound waves of varying intensity and pitch. Your behaviour can also be seen in a wider context: it is just one aspect of what is involved in studying psychology. Take a few minutes to jot down your explanation for your behaviour: if someone were to ask why you are behaving in the way you are, what would your answers be? Try to think of many different ways of answering the question. List too any processes that you think might be going on in your mind – how would you describe them?

## COMMENT

The first thing to note is that your behaviour can be explained in many different ways. For example, you might have noted that your reading is bound up with a *feeling of elation* – perhaps you love studying cognitive psychology – or a *feeling of anxiety* – perhaps you are uncertain of obtaining a good course grade. Your explanation adverts to *emotions*. Perhaps you jotted down as an answer that you *reasoned* that you ought to read this book since you want to do well on your course. Perhaps doing well on your course is part of a strategy to reach a goal, or *solve a problem* such as how to improve your qualifications. You might also have suggested that you *decided* to read this book – perhaps faced with different ways of spending your time, you judged that this would be the most beneficial (we'll try not to let you down!). You might have thought there are processes going on in your mind to do with *reasoning, problem solving* and *decision making*.

It might be that you are reading this chapter for a second time because you want to make sure you *remember* it. So, your explanation adverts to *memory*, and the processes that are responsible for things being remembered (and forgotten).

How else might you have explained your behaviour? You might have suggested that you were trying to *understand* the chapter; that you behaved the way you did because

you were involved in understanding words, phrases, and sentences. You may have indicated that there must be processes for understanding *language*. Perhaps there were other explanations you offered. Maybe you explained your reading of the book by saying 'That is what books are for' – because you categorized it as a book. Maybe you suggested you were scanning your eyes across the page in order to perceive and recognize words. And, just maybe, you suggested that your behaviour was happening because you were paying *attention*, and not being distracted by a telephone or a door bell.

The words in emphasis in the previous paragraphs all provide important means for explaining behaviour that are used by cognitive psychologists, and are all major topics of this book.

---

Activity 1.2 shows how everyday behaviour can be explained in a number of different ways, and as involving many different kinds of cognitive process. In fact, all of the types of explanation referred to in the comment on Activity 1.2 are ones that will be developed at some length in this book. However, a corollary of the observations made in Activity 1.2 is that cognitive psychologists try to devise studies that isolate the particular cognitive processes under investigation – for example, a researcher interested in language processing will try to devise their studies so that they measure language processes only, and are not unwittingly influenced by other processes, such as emotion or reasoning. Consider also how the studies referred to in Boxes 1.1 to 1.3 try to focus exclusively on the issue of category specificity. Indeed, it is a general strategy within cognitive psychology to try to isolate particular cognitive processes for further investigation. Table 1.1 lists some prevalent assumptions to which this strategy gives rise.

**Table 1.1** Assumptions commonly made in the cognitive approach

1	It is assumed that cognitive capacities can be partitioned such that individual capacities can be studied in isolation (e.g. so that language can be studied in isolation from memory)
2	Cognitive psychology tends to focus on the individual and their natural environment (relatively de-emphasizing the roles of culture and society)
3	Cognitive capacities are assumed to be relatively autonomous from non-cognitive capacities (e.g. affect, motivation, etc.)
4	It is assumed that it is useful (and meaningful) to distinguish 'normal' from 'abnormal' cognition
5	Adults are assumed to be sufficiently alike that we can talk of a 'typical' cognizer, and generalize across cognizers, ignoring individual differences
6	Answers to basic, empirical questions can be given in terms of information processing
7	Answers to basic, empirical questions should be justified on empirical grounds
8	Answers to the basic, empirical questions must be constrained by the findings of neuroscience (as and when these are relevant)

Source: adapted from Von Eckardt, 1993, pp.54–5

## Summary of Section 2

---

- Cognitive psychology can be characterized as the scientific study of the mind.
- Cognitive psychology can be characterized in terms of its methods:
  - experimental studies of normal cognition
  - neuropsychological studies that relate normal to abnormal cognition
  - neuroimaging studies that reveal the location and/or the time course of brain activity
  - computational models which can be tested and compared with experimental data.
- Cognitive psychology can be characterized in terms of its subject matter (see the table of contents for this book).
- Everyday behaviour involves multiple cognitive processes:
  - cognitive studies tend to isolate one process or set of processes for study.

## 3 A brief history of cognitive psychology

Cognitive psychology did not begin at any one defining moment, and there are many antecedents to its evolution as a branch of enquiry. In this section we will briefly sketch some of those antecedents and try to indicate how and why they resulted in the development of what today we call cognitive psychology. However, all written history is necessarily selective and simplified, and a historical account as brief as the one we are about to give must be especially so. We start with introspectionism.

### 3.1 Introspectionism

Modern experimental psychology has its roots in the work conducted in Europe in the mid nineteenth century by such people as Donders, Fechner, Helmholtz and Mach. When Wundt established the first dedicated psychology laboratory in Liepzig in 1879, he sought to build upon the efforts of these pioneers. He took *consciousness* to be the proper subject matter of psychology. According to Wundt, physical scientists study the objects of the physical world either directly or, more often, through observation of the readings on instruments. In either case, observation is mediated by conscious experience, but for physical scientists things in the world are the object of study not the conscious experience by means of which we know them. Psychology would be different in that it would take as its subject matter conscious experience itself.

Wundt adopted **introspection** as a research method, believing that properly trained psychologists should be able to make observations of their own experience in a manner similar to the way properly trained physicists make selective observations of the world. Wundt fully understood the need to design experiments with adequate controls and to produce replicable results. He also made use of objective measures of performance, such as reaction time (RT). The focus of his interest, however, was the

conscious experience that preceded the response. For example, if one condition in an experiment yielded longer RTs than another, he wanted to know how the two preceding conscious experiences differed. Wundt was not concerned with the unconscious processes involved in responding to a simple stimulus – the rapid information-processing operations that, as you will find in the following chapters, form much of the subject matter of modern cognitive psychology. He considered these to lie in the realm of physiology rather than of psychology.

In opposition to Wundt's Leipzig school was the Würzburg school of introspection. Its leader, Külpe, was a former student of Wundt's, who with his colleagues and students developed an alternative view of conscious experience and what could be revealed by introspection. We can characterize the main difference between the two schools in terms of a distinction that will be more fully introduced in Chapter 3 in relation to the topic of perception, although the protagonists would not have used these exact terms themselves. Put simply, the Leipzig school held that the contents of consciousness are constructed 'bottom-up' from simple sensations combined in accordance with the strength of association between them (something like the connectionism you can read about in Chapters 4, 16 and 17). The Würzburg school, on the other hand, held that the contents of consciousness are determined in a much more 'top-down' fashion by the nature of the task that one is engaged upon. Külpe and his colleagues sometimes studied simple tasks, but tended to favour more complex ones in which mental acts such as attending, recognizing, discriminating and willing played a larger role.

Introspectionism went into a terminal decline during the first two decades of the twentieth century. The details of the many unresolved disagreements between the two schools of introspectionism need not detain us here, but it is worth noting two things. First, the introspectionists developed elaborate classifications of conscious experience, a topic that has quite recently begun to attract the attention of psychologists once again (see Chapter 15). Second, although psychologists began to lose interest in consciousness during those two decades, the exploration of consciousness still remained central to developments in the visual and literary arts (e.g. cubism and expressionism in painting, and James Joyce, Virginia Woolf and Gertrude Stein in literature).

### 3.2 Gestalt psychology

The perceived failures of introspectionism provoked a number of intellectual reactions. In Europe, the gestalt psychologists built upon the work of the Würzburg school and argued that the contents of consciousness cannot be analysed into simple component sensations. According to Wundt, the perception of movement results from a sequence of sensations corresponding to an object occupying successive locations over time. However, Wertheimer argued in 1912 that 'pure movement' can be perceived directly; it does not have to be 'inferred' from changes in the location of an object. A good example is when we see the wind gust through grass. Blades of grass bend in succession but no blade changes location. What we perceive is pure motion (of the invisible wind) without a moving object. (Modern studies show that motion perception can, in fact, arise either on the basis of the changing location of an object or from successive changes across space without a moving object.) Gestalt psychologists also emphasized the importance of the perception of stimulus

patterning to our conscious experience. A tune played in one key on one sort of instrument remains the same tune when played in another key or on a different instrument. Since the notes, or the sounds making up the notes, have changed in each case, there must be more to the tune than can be found by an analysis into simple auditory sensations. The tune is in the perceived relationships between the notes, their patterning.

Meanwhile, in the USA, William James opposed introspectionism with his ‘functionalist psychology’. Sounding remarkably like an exponent of what is now called evolutionary psychology, James stated that, ‘Our various ways of feeling and thinking have grown to be what they are because of their utility in shaping our reactions to the outer world’. These functions of the mind were, in James’s view, the proper subject matter for psychology. Perceiving and thinking, grief and religious experience, as psychological functions, were themselves to be the focus of interest, rather than the evanescent contents of consciousness on which the introspectionists had fixated. However, James’s ideas were soon to be largely swept aside by another and more powerful current in US thought, which was behaviourism.

### 3.3 Behaviourism

The founders of behaviourism were driven by various motives, not all shared in common. Watson, the principal standard-bearer for the new kind of psychology, was especially keen to move psychological research out of the laboratory and into ‘the real world’. He was less interested in fine distinctions of conscious experience than in how people act in everyday life, and in how they can be influenced. He wanted to see psychological knowledge applied to education, clinical problems and advertising, and he initiated work in all these areas. Not all behaviourists were as zealous as Watson when it came to applying psychology, but one belief they did have in common was that psychology should be scientific and objective; and by this they meant that its subject matter should be publicly observable. Consciousness is (at best) only privately observable; it is not publicly observable. What is publicly observable is behaviour and stimuli. So psychologists such as Thorndike, Watson and, later, Skinner, Eysenck and others argued that psychology should be scientific in its approach, and should seek to explain behaviour through reference only to stimuli. The emphasis on public observation was intended to place psychology on an objective footing, akin to the natural sciences like physics and chemistry, and it reflected a wider philosophical consensus as to the proper nature of scientific enquiry.

#### 3.3.1 Science and the unobservable

In all human efforts to comprehend the world there is a tension between, on the one hand, observable events and, on the other hand, the often encountered need when explaining them to postulate unobservable theoretical entities and forces, whether gods or atoms. This tension is central to science. A key idea in the development of science has been that knowledge should be empirical, based on experience not on received wisdom or purely rational calculation. Observation is one of the touchstones of science, but scientific theories also refer to unobservables. The explanation that physics offers for an apple falling to Earth invokes the notion of a gravitational force, something that is not directly observable. Similarly, in

explaining why a compass needle points to magnetic north, physicists talk of magnetic fields, and lines of magnetic force. But these things too are unobservable. If you have ever placed iron filings near a magnet, you will see that they will move to orient themselves along the lines of the magnetic field. But, strictly, we don't observe the magnetic field, nor the lines of magnetic force, but rather their influence upon the iron filings. All natural sciences employ unobservable, theoretical constructs that are invoked in order to explain observations. For example, chemistry appeals to notions such as the energy levels of electrons in order to explain why compounds react. These levels are unobservable too, of course. So, the fact that a discipline is committed to explaining observed behaviour by reference to hypothesized, unobservable constructs does not in itself render the discipline unscientific.

But to find scientific acceptance, unobservable constructs have to be seen to do useful theoretical work. When Newton proposed the notion of a gravitational force, certain critics immediately accused him of introducing a mystical notion into 'the new science'. Newton's ideas gained acceptance only because they met other scientific criteria – such as elegance, simplicity and rigour – and because the concept of gravitation, despite its somewhat mysterious nature, had a wide range of application. Gravitation explained not just the fall of objects to the ground but also the rhythm of the tides and the movements of the planets. It could also be precisely formulated mathematically as an inverse square law: the attraction between any two bodies varies as the square of the distance between them. In other words, the willingness of the scientific community to countenance a hypothetical unobservable depends on how useful it is judged to be on a range of criteria.

Science has had to live with the necessity for unobservables. But acceptance through necessity is not liking, and science always receives a boost when a technical breakthrough for the first time brings a previously unobserved entity into the realm of observation. For example, Mendel postulated 'units of heredity' on the basis of his plant-breeding observations, but these ideas were felt to be on a firmer footing once new technology made it possible to see chromosomes and genes. Thus, scientists are forced somewhat grudgingly to accept the need for postulating unobservables. And because science – like all human institutions – is subject to swings of fashion, the willingness to countenance unobservable theoretical entities fluctuates over time. For reasons which we are unable to describe here, but which were rooted in the growing crisis of classical physics that would culminate in the birth of quantum theory and relativity theory, the late nineteenth and early twentieth century was a period during which scientists were particularly intolerant of unobservables. The importance of observation became enshrined in the assumption known as **operationism**. This is the idea that theoretical concepts are only meaningful to the extent that they can be exhaustively analysed in terms of things that can be observed.

### 3.3.2 Back to behaviourism

The bias against unobservables affected all the traditional sciences and also the newer, aspirant scientific disciplines such as physiology and psychology. The introspectionists, with their 'observations' of consciousness, had responded to it, but the intellectual climate seems to have been especially suited to propagating an emphasis on what could be publicly observed. With the decline of introspectionism, behaviourism was taken up enthusiastically, first in the USA and then more widely.

While behaviourists could, perhaps, concede the *existence* of consciousness while arguing that it was not appropriate for scientific study, at least some of them felt that operationism committed them to the stronger claim that talk of consciousness was not even meaningful. Of course, behaviourism has never been a single view, and since the time of Watson and Thorndike behaviourists of various hue have modified their positions. Skinner, for example, conceded that internal mental events, including conscious experiences, might *exist* (indeed they were construed as forms of covert behaviour). But despite this rejection of operationism, even Skinner still thought that talk of internal events should be avoided within a scientific psychology.

You might think that avoiding talk of internal events might make it impossible to explain many, or even most, psychological phenomena. However, behaviourists were concerned to show how even complex phenomena might be understood in terms of principles of learning, with behaviour seen as made up of learned responses to particular stimuli. One view of language production, for example, was that the utterance of a word could be seen as a learned response. The utterance of a whole sentence could be seen as involving a chain of stimulus–response pairs, in which each response (the utterance of a word) also serves as the stimulus that leads to the production of the next response (the next word).

Despite the possibility of giving behaviourist explanations of complex activities such as the utterance of a sentence, behaviourists tended not to offer accounts of what we now refer to as higher mental processes – processes such as producing and understanding language, planning, problem solving, remembering, paying attention, consciousness and so on. As the years passed, however, some psychologists came to see this as a major failing.

### 3.4 The return of the cognitive

In 1948, at a meeting known as the Hixon symposium, Karl Lashley gave a talk entitled ‘The problem of serial order in behaviour’ (Lashley, 1951). In this, he gave prominence to the problems posed for behaviourist accounts by complex actions in which behaviour segments are somehow linked together in a sequence, and where two segments depend upon one another, even though they may be separated by many intervening segments. Language, as you might have guessed, provides a prime example. In fact, the last sentence illustrates the point nicely: when I came to write the word ‘provides’ in the previous sentence I chose to end it with the letter ‘s’. I did so, of course, because this verb has to agree grammatically with the singular noun ‘language’, the subject of the sentence. In my actual sentence, these two words were separated by a clause, and so my action at the time of writing the word ‘provides’ depended upon a much earlier behaviour segment – my writing of the word ‘language’. Lashley argued that since the production of some words in a sequence could be shown to depend upon words produced much earlier, the simple view that each word is the stimulus that produces the subsequent word as a response could not properly explain language production.

He also argued that many behaviour sequences are executed simply too rapidly for feedback from one segment to serve as the trigger for the next. He cited examples such as the speed with which pianists and typists sometimes move their fingers, or with which tennis players adjust their whole posture in response to an incoming fast

service. Lashley's alternative to the chaining of behaviour segments was to suppose that complex sequences are planned and organized in advance of being initiated. The speech errors discussed in Chapter 7 of this book provide especially compelling examples of the kind of planning and organization that underlie skilled behaviour.

Lashley's view that behaviourism could not properly explain how people produce (or comprehend) language was later reinforced by a review of Skinner's book *Verbal Behavior* (1957) by the linguist Noam Chomsky (1959). Chomsky argued, contra behaviourism, that language could not be thought of as a set of learned responses to a set of stimulus events. His argument had a number of different aspects. For example, he argued that children seem to acquire their first language too effortlessly – if you have tried to learn a second language you can perhaps testify to the difference between learning a first and learning a second language. While the latter seems to require intensive and effortful study, the former is something that pretty much everyone does without the need for formal schooling. He also argued that if the behaviourists were right, then exposing children to impoverished or ungrammatical language should hinder their learning of the correct stimulus–response relationships. Yet studies show that much of the speech to which young children are exposed is indeed ungrammatical and otherwise impoverished, and this in no way prevents them from learning the grammar of their native tongue. Similarly, he argued that general intelligence ought to influence the learning of stimulus–response relationships. Again, however, intelligence does not seem to influence whether or not children learn the underlying grammatical rules of their language. Chomsky presented many other arguments to the same effect, and though many of these have been thought to be contentious, his position was extremely influential in setting up an alternative, cognitive conception of language. Most significantly, Chomsky proposed that language is rule-based and that, far from children learning language by learning how to respond to particular stimuli, their acquisition of language involves acquiring its rule-base. On this view, my being able to write grammatical sentences involves deploying my (generally implicit, or unconscious) knowledge of the rules of language. In referring to such implicit knowledge, Chomsky proposed that an understanding of how people produce, comprehend or acquire language will necessarily involve reference to something that cannot be directly observed – their knowledge of the underlying rules, or organization, of the language.

Although this emphasis on the role of planning, organization and rules in the generation of behaviour was to be hugely influential from the 1950s onwards, these ideas were certainly not new to psychology. As mentioned previously, the gestalt psychologists had drawn attention earlier in the century to the importance of patterning, or organization, for perception, and the same point was also made in relation to action. Someone who has learned to sing or hum a tune can very probably manage to whistle it thereafter. Yet singing, humming and whistling call for very different sequences of muscle movements. This indicates that learning a tune must involve learning a set of abstract relationships between notes which can be instantiated as any of a variety of muscular productions. A similar idea, that what is learned must often be more abstract than straightforward stimulus–response connections, was also expressed by the school of ‘cognitive behaviourists’ associated with Tolman (1932). Rats that had learned, for example, repeatedly to

turn left in a maze to find food were shown to swim left when the maze was flooded. Since the muscle movements of running and swimming are completely different from one another, the rats must clearly have learned something more abstract than a particular chain of muscular responses.

Even before the writings of the gestalt psychologists or the work of Tolman, psychologists studying the acquisition of skills had realized the importance of planning and organization for the production of skilled behaviour, such as in Morse telegraphy or typing (Bryan and Harter, 1899). At the time of the Hixon symposium, therefore, there were already existing traditions within psychology upon which the renewed interest in the planning and structure of behaviour could draw. And, of course, the intellectual climate of the mid twentieth century was changing rapidly in many other ways too. New technologies were influencing the ability of scientists to conceptualize the workings of complex systems. One of the most crucial issues related to the type of causal explanation that is appropriate to explain the behaviour of such a system. Purposive, or teleological, explanations had been taboo in Western science since the time of thinkers such as Galileo and Newton. Where, for example, an ancient Greek philosopher might have said that a stone falls to earth ‘in order to’ reach its natural resting place at the centre of the earth (which was also the centre of the Greek universe), Newton said that the stone falls because it is acted upon by the force of gravity. The strategy of explaining phenomena in terms of causes that precede and ‘push’ their effects, rather than in terms of goals, or final states, towards which events are ‘pulled’, had proved highly successful in the physical sciences. The move from goal-directed, purposive explanations to mechanical cause-effect explanations was usually considered to be a move from prescientific, animistic thinking to proper scientific thinking. Behaviourism was, and still is, an attempt to bring psychology into step with this way of analysing phenomena. A strict emphasis on an organism’s history of conditioning allows an explanation of behaviour in terms of prior causes rather than of future goals. However, the development of progressively more complex artificial devices started to call into question the universal applicability of explanations in terms only of prior causes. It became increasingly clear that, while the functioning of the mechanical parts of any such system *can* be explained in cause-effect terms, such explanations will never capture the function (or purpose) of the whole system.

Central to the new kind of apparently purposive machines (known as servomechanisms) was a reliance on feedback loops. **Feedback** is information about the match or mismatch between a desired goal-state and an existing state of affairs. The classic example is the domestic central heating system, in which the thermostat setting selected by the householder is the goal-state and the temperature measured by an air thermometer is the existing state. The two are compared mechanically. If the existing temperature is less than the desired temperature, this negative feedback is transmitted to the boiler controls causing the boiler to be switched on. The boiler continues to fire until information has been fed back to the boiler controls that the discrepancy between the actual and desired temperatures has been eliminated. The system as a whole exhibits a simple but dynamic behaviour, with the boiler turning on and off in a manner that maintains room temperature at or about the desired level. Importantly, the function of maintaining a steady temperature cannot be localized to any one component of the heating system, such

as the thermostat, the thermometer, the boiler or its controls, but is a property of the system – as a whole.

Far more complicated servomechanisms with more complex feedback controls were also being developed. Anti-aircraft gunnery may not seem very pertinent to an understanding of animal and human behaviour, but it was partly as a result of working on gunnery problems in the Second World War that the mathematician Norbert Weiner developed the notion of ‘cybernetics’, the science of self-governing, or goal-directed, systems. Accurate anti-aircraft gunnery requires that a projectile is fired, and timed to explode, not at the present location of the target aircraft but at its future location. This means not only predicting the future position of the plane but also rotating the gun so it faces in the appropriate direction and with the correct elevation. Clearly, humans successfully extrapolate flight paths and aim at future positions when, for example, shooting game birds. However, for planes flying at ever greater heights and speeds, calculation of the necessary trajectory of the projectile exceeds human capabilities and must be computed automatically. Moreover, using motors to move a gun weighing many tons is a very different matter from moving a shotgun, or indeed a bow and arrow, held in your arms. Although we are mostly unconscious of it, normal bodily movement is based upon continuous muscle, tendon and visual feedback about how the movement is proceeding. Unless similar feedback is designed into the gun control system, the swinging anti-aircraft gun may easily undershoot or overshoot the intended position, particularly as, depending on the air temperature, the grease packed round the mechanism will be more or less ‘stiff’. Apply too little power and the gun will undershoot the intended position, a second push will be required and the gun will ‘stutter’ towards its position. Apply too much force and the gun will overshoot, and will have to be pulled back, in what can turn into a series of increasingly wild oscillations. Engineers discovered that the smoothest performance was achieved by using feedback loops to dynamically control the turning force applied to the gun.

Weiner, and other cyberneticists such as Ashby, recognized the importance of feedback and self-correction in the functioning of these new and complex technological devices, and they also saw analogies with complex natural systems. Weiner drew parallels between the effects of certain neurological conditions and damage to the feedback control of behaviour. For example, the tremors observed in Parkinsonian patients were likened to the oscillations of an anti-aircraft gun when its movement is insufficiently ‘damped’ by feedback control.

An important intellectual leap for cognitive psychology came with the realization that just the same kind of analysis can be applied at any level of behavioural control. In other words, it is not just automatic homeostatic functions or unconsciously executed movements that can be analysed in terms of feedback loops but any function/behaviour from the wholly non-conscious to the fully conscious and intended. Miller *et al.* (1960) developed the notion of feedback control into the hypothesis that behaviour (of animals, humans or machines) can be analysed into what they called **TOTE units**. TOTE stands for Test-Operate-Test-Exit. A test is a comparison between a current state and a goal-state. If a discrepancy is registered, some relevant operation intended to reduce the discrepancy will be performed (e.g. switch on the boiler). A second test, or comparison, is then conducted. If a

discrepancy remains, the operation can be repeated, followed by another test. If the discrepancy has been eliminated, the system exits the TOTE unit.

Miller *et al.* conceived of the TOTE unit as an advance on the conditioned reflex notion of Pavlov and the conditioned response notion of Watson and Skinner, both of which can be conceptualized as TOTEs. The aim was to develop a unit of analysis of behaviour that could apply to everything from a dog's conditioned salivatory response to deliberate, planned action. The TOTE provides a basic pattern in which plans are cast; the test phase specifies what knowledge is necessary for a comparison to be made, and the operation phase specifies what the organism does about the outcome of the comparison. Although this scheme makes it possible to talk about purposive behaviour, and about unobservable goals and comparison operations, there is continuity from behaviourism. Cognitive psychology generally attempts to retain the scientific rigour of behaviourism while at the same time escaping from the behaviouristic restrictions in relation to unobservables.

An important property of TOTEs is that they can be nested within hierarchies. The operation segment of any TOTE can itself be composed of one or more TOTE units. For example, the TOTE for starting the car might be nested within the operation of a larger TOTE for driving to the shops, which might itself be nested within a still larger unit having the goal of buying a present. This nesting of feedback loop units provides a way to conceptualize how behaviour can be complexly structured. In this scheme, moment-to-moment control of behaviour passes in sequence between a series of TOTE goal-states, with the TOTE units themselves nested in hierarchies. Miller *et al.* explicitly likened this 'flow of control' of behaviour to the way in which control in a computer program switches in orderly fashion from command line to command line as the execution of any particular subroutine is completed. (Note: what 'flows' around a TOTE can be energy, information or, at the highest level of conceptual abstraction, control.)

### 3.4.1 Computers and the mind

Another development in the mid twentieth century with a huge import for the development of cognitive psychology was the opening up of a new field concerned with the possibility of designing and then building computers. Building on earlier work that developed a formal, or mathematical approach to logical reasoning, Claude Shannon in 1938 showed how core aspects of reasoning could be implemented in simple electrical circuits. In the 1940s, McCulloch and Pitts showed how it was possible to model the behaviour of simple (and idealized) neurons in terms of logic. Taken together, these developments suggested something that at the time seemed extraordinary – that the brain's activity could, at least in principle, be implemented by simple electrical circuits.

In parallel with these developments, the 1930s and 1940s saw pioneering theoretical developments in computation and information processing. Turing, in 1936, developed an abstract specification for a machine (a Turing machine) that could compute any function that in principle could be computed. In the 1940s, Shannon and Weaver used the tools of mathematics to propose a formal account of information, and of how it could be transmitted.

Technological progress was also rapid. In 1941, Konrad Zuse of Berlin developed the world's first programmable, general-purpose computer. In 1943,

Colossus, a special-purpose computer designed to break wartime codes, became operational at Bletchley Park, in Buckinghamshire. In 1946, John von Neumann articulated a set of architectural proposals for designing programmable, general-purpose computers. These were adopted almost universally and computers have since also been known as von Neumann machines. In 1948, the Manchester University Mark I programmable, general-purpose computer became operational and, in 1951, Ferranti Ltd began producing, selling and installing versions of the Manchester Mark I – the world's first commercially available, programmable, general-purpose computer.

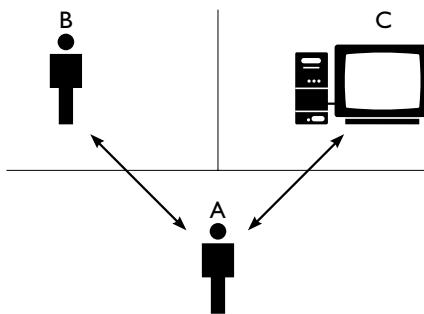
These developments, fascinating though they were in their own right, also seemed to carry important implications for our understanding and study of the mind. They appeared to show, for instance, that reasoning, a central feature of the human mind, could be implemented in a digital computer. If that were the case, then not only could the computer be used as a tool to aid our understanding of the mind, but the question would also arise as to whether minds and computers are essentially alike. Indeed, in 1950, Turing proposed a test – the Turing test – by which he thought we should judge whether two entities have the same intelligence. Turing believed that, should the situation ever arise whereby we could not distinguish the intelligence of a human from the ‘intelligence’ of a computer, then we ought to concede that both were *equally* intelligent. Moreover, since we are in agreement that humans are capable of thought, we also ought to concede that computers are also capable of thought! Box 1.4 (overleaf) outlines the Turing test and considers what it might take for it to be passed.

Turing's position remains controversial, of course, though it certainly captured the imagination of the time. In 1956, at the Dartmouth Conference (held in Dartmouth, New Hampshire), John McCarthy coined the phrase ‘Artificial Intelligence’ (or AI). He founded AI labs at MIT in 1957, and then at Stanford in 1963, and so began a new academic discipline, predicated on the possibility that humans are not the only ones capable of exhibiting human-like intelligence.

You have now been introduced to a variety of the influences that go to make up cognitive psychology. Cognitive psychology inherits some of the behaviourist concerns with scientific method. Throughout this book you will see that almost constant reference is made to systematic observations of human behaviour (and sometimes animal behaviour too). Almost every chapter will present the results of empirical investigations, and these are fundamental in guiding our understanding. But cognitive psychology rejects the exclusive focus on what is observable. As Chomsky implied, understanding the mind requires us to consider what lies behind behaviour – to ask what rules or processes govern the behaviour we observe. Each chapter will also consider the extent to which we understand how the mind processes information, and how that information is represented. Cognitive psychology also has a major commitment to the use of computers as a device for aiding our understanding of the mind. First, computers are used as research equipment to control experiments, to present stimuli, to record responses and to tabulate and analyse data. Second, computers are also used as a research tool – if we can implement reasoning in a computer, for example, we may gain insight into how reasoning might be implemented in the brain. So, most of the

**1.4****The Turing test: can computers think?**

Turing proposed that we could determine whether a computer can think by judging whether it succeeds in what he called the imitation game. In the game there are three participants, two humans (*A* and *B*) and a computer (*C*). The arrangement of the participants and the communication flow between them is schematically indicated in Figure 1.3.



**Figure 1.3** The arrangement of the participants in Turing's imitation game

The participants are positioned in separate rooms, so each one is unable to see, hear or touch the others. However, one of the human participants (*A*) is connected via a VDU terminal connection to the other human participant (*B*) and also to the computer (*C*). *A* can communicate electronically with both *B* and *C*. The goal for *A* is to ascertain which of *B* and *C* is the computer, and which the human. The goal of *B*, the other human, is to assist *A* in making the correct

identification (perhaps by trying to appear as human as possible). *C*'s goal, by contrast, is to lead *A* into making the wrong identification (by imitating human behaviour). *C* wins the game if *A* cannot reliably identify *C* as the computer. Turing's claim was that if a computer could simulate human behaviour so successfully that another human could not tell that it was a computer, then the computer could legitimately be said to think.

chapters in this book will also discuss ways in which researchers have used computer models to help us understand how the mind processes and represents information when people perform certain behaviours. Third, and more controversially, computers are also considered to be candidate 'thinkers' in their own right. Understanding more about the nature of computation itself may shed light on the nature of thinking, and on the nature of the mind.

## Summary of Section 3

- Cognitive psychology inherits some of the behaviourist concerns with scientific method. Almost every chapter in this book presents the results of empirical investigations, investigations that are fundamental in guiding our understanding.
- Cognitive psychology rejects an exclusive focus on what is observable. Almost every chapter considers the extent to which we understand how the mind processes information, and how that information is represented.
- Cognitive psychology is committed to using computers as a tool for aiding our understanding of the mind.

- Introspectionist and gestaltist interest in conscious experience was replaced by the behaviourist focus on what is publicly observable.
- There is always a tension in science between the emphasis on observation and the need to postulate unobservable theoretical entities.
- Behaviourists did not necessarily deny the importance of higher mental functions, but rarely offered accounts of them.
- Cognitive psychology has many roots; it has been heavily influenced by technological developments and the way they help us to understand complex behaviours.

## 4 Science, models and the mind

If cognitive psychology is concerned with the processes and representations of the mind, and these cannot be directly observed, how can cognitive psychologists bridge the gap? How do we speculate about the nature of something we cannot observe, while remaining scientific? There are broadly three kinds of answer.

First, as we have already discussed, scientific theories commonly invoke unobservable theoretical entities to account for observational data (e.g. force fields, electron energy levels, genes or cognitive operations).

The second answer builds on the first. When a theory hypothesizes an unobservable, theoretical construct, a model needs to be specified of the relationship between the construct and the behaviour to be explained. It would have been insufficient for Newton to have tried to explain why things fall to Earth by simply invoking the notion of gravitation. He went further and derived equations to model the effects of gravity, which can be used to generate predictions about how gravity ought to work for things whose motion has not yet been systematically observed. So physicists could then perform studies in order to confirm the predictions (that is, until Einstein's theories of relativity, but that is another story).

Cognitive psychology proceeds in a similar way. Consider again the example of language. Cognitive psychologists have made numerous detailed observations of the production (and comprehension) of language (you can find discussions of these in Chapters 6 and 7). Explaining these observations, however, seems to require positing things internal to the mind that are involved in producing the observed behaviour. These are the unobservable, theoretical constructs of mental processes and structures. Positing these, of course, is just the starting point. The challenge for cognitive psychologists has been to say more. They have to develop models of these mental structures and processes, show how they give rise to the observed behaviour, and, importantly, show how successfully they predict behaviour that has not yet been systematically studied in experiments.

Developing a model is not easy; Newton apparently needed the inspiration provided by an apple falling to Earth (or so the story goes). And much of the challenge facing cognitive psychologists is to harness their creativity and imagination in order to suggest plausible models. Throughout your reading of this book, you might wish to consider how you would have responded to some of the

problems described. You might want to consider what would constrain your choice of model, what kinds of model you would have developed, and how you would have set about doing this. Without doubt, these are difficult questions – so don't lose too much sleep over them! – but they at least serve to show how creative cognitive psychology is. Creative too is the matter of devising studies in order to evaluate a model. By working out the predictions a model might make, psychologists can evaluate it by devising studies to test its predictions, and by then making the relevant behavioural observations.

Creating models and designing studies to test them is not easy, but cognitive psychologists can use computers to help. The previous section suggested two ways in which computers are important to cognitive psychology other than as experimental equipment – computers might be capable of thought; and they can also serve as tools for implementing models such as a model of language processes. Now, perhaps, you can see how they might contribute to the scientific objectives of cognitive psychology – researchers can use computers in order to create models. Just as computer programmers can build programs to do things such as word processing, or financial accounts, so researchers in cognitive psychology can program computers to behave according to a particular model of the mind. Using computers to program particular models can be helpful on a number of counts:

- 1 Models can rapidly become very complicated – too complicated to be expressed verbally, or for one person to hold all the relevant details in mind. This problem affects others too – meteorologists increasingly use computer models of weather systems, and economists use computer models of the economy. The phenomena involved are so complicated that, without computers, they would be almost impossible to model.
- 2 It is not always easy to work out the predictions of a model. Programming a model can allow researchers to simulate the effects of different conditions and so find out how the model behaves, and whether this behaviour accurately predicts how humans will behave.
- 3 Perhaps most important of all, by programming a model into a computer researchers can determine whether the model is internally consistent (whether there are statements in the model that contradict one another), and whether the model is already clearly and precisely stated. If it is, the computer program will run; otherwise, it will crash.

So cognitive psychology can posit the existence of unobservable (cognitive) processes and structures and still be scientific. Not only is this true of other disciplines like physics and chemistry, but, like those disciplines, the gap between observable behaviour and unobservable processes and structures can be bridged via the creation and evaluation of models.

There is, however, a new possibility for linking cognitive processes with a focus on observation, and this leads to the third answer to the question with which this section began. The advent of new techniques for imaging the brain suggests that, just possibly, mental processes and structures may not be entirely unobservable (as the behaviourists once believed).

Functional MRI studies (and other kinds of imaging) allow us to see which parts of the brain become especially active when people are engaged in a certain task (relative to when they are engaged in some control task or tasks). There is considerable debate in the cognitive community as to the usefulness of imaging techniques for helping researchers to develop theories of cognition. Activity 1.3 will help you get a sense of the issues involved.

### ACTIVITY 1.3

---

Consider again the brain images in colour Plates 1 and 2. First, think about what you could infer from the images alone. What does the indication of activity in particular brain regions tell you? Second, think about the processes going on inside participants' minds. What additional information would you need to be able to say what the brain activity represents? Suppose you were given very detailed anatomical descriptions of the active regions: what would that enable you to conclude?

### COMMENT

It is one thing to say that there is activity in particular regions of the brain, yet quite another to say exactly what cognitive processes and structures are involved. An image of brain activity, on its own, does not help very much. Seemingly, what is crucially needed is further information as to what information each brain region processes. That is, we need to know the function of the active regions. One way of trying to identify the function of different brain regions is to compare brain images for different kinds of task – regions that are active for all tasks may be implicated in information processing that is common to those tasks. This assumes we have good models for the information-processing characteristics of different tasks. If so, and also using anatomical and neuropsychological evidence, researchers can then tentatively begin to identify particular regions with particular functions. This in turn can help researchers to interpret and design further brain-imaging studies.

---

One criticism of imaging studies is that, at best, they help researchers to localize a particular function – that is, researchers can identify the function with a particular region of the brain – but that they do not improve our theories of cognition. However, this is a bit like saying that being able to see chromosomes and genes down a microscope does not improve the theory of genetic inheritance. In one sense that is true, but making visible entities that were previously only theoretical does increase overall confidence in the theory. Similarly, suppose a cognitive theory says that reading some words involves using a visual processing route and reading other words involves using an auditory processing route. Finding that the first task induced activity in areas known to be engaged by other visual tasks, and that the second task induced activity in areas known to be engaged by auditory tasks would increase our confidence in the theory.

Without prejudging the ongoing debate in this area, it is likely that imaging techniques will contribute to cognitive theory in various ways. Sometimes the contribution will be at the level of theoretical deduction, sometimes it may be at a less palpable level as when it adds to the confidence in a theory. When genes were

first made visible, genetic engineering was a very distant prospect, but it is hard to imagine the latter without the former. The advances in cognitive sciences to which neuroimaging will contribute are equally hard to predict, but we shall be surprised if they do not prove to be many and varied.

## Summary of Section 4

---

- Cognitive psychology can be scientific, while being interested in what goes on, unseen, inside the mind, for a number of reasons:
  - other natural sciences invoke unobservable entities and are not as a consequence rendered unscientific
  - like other sciences, cognitive psychology proceeds by modelling unobservables to produce predictions which can be tested by conducting appropriate studies
  - the advent of brain-imaging technology, though undoubtedly contentious, raises the prospect of observing processes that were previously unobservable.

## 5 The cognitive approach

Thus far, we have talked of cognitive structures and cognitive processes. Section 3 offered some examples of historical proposals as to what kinds of things cognitive structures and processes are. Contemporary cognitive psychology equates representations with cognitive structures, and computations over these with cognitive processes.

### 5.1 Representation

We have emphasized the scientific nature of cognitive psychology. However, Fodor (1974) argued that psychology might be a special science – special because its subject matter, the mind, stands in a complex relation to the material, physical world – and therefore takes a different form from the natural or social sciences. Spelling out the relationship between the mind and the physical world, even between the mind and the body, is extremely difficult. Two competing intuitions have guided people's thinking about the issue. One is that the mind transcends the physical body (and the brain) – that when we say we are in love, for example, we mean more than that we are in a particular bodily or brain state. Though you may share this intuition, it is difficult indeed to say what a psychological state is if it is *not* physical. It is also difficult to reconcile this intuition with the methods of natural science – how is it possible to study something scientifically if it is not physical in nature? The competing intuition is that all aspects of humanity, including our minds, ought to be explicable as parts of the natural world, and so explicable by the natural sciences. Humans are, after all, products of natural, evolutionary pressures, shaped by the world in which we have

evolved. How could we come to possess a mind that could not be explained as part of the natural, physical world?

The tension between these two intuitions is real and difficult to resolve (as you will see from Chapters 15 and 17). Here we can do no more than hint at the difficulties. One feature of the mind may go some way to showing why the intuitions are so difficult to reconcile. It is the feature of representation.

Some things in the world have the property of being ‘about’ something else. Books, for example, tend to be about other things. A book on the Second World War is about precisely that – the real events that go to make up the Second World War. The observation is so mundane that you may never have given it a second thought. Yet this property of **aboutness** is quite extraordinary, and certainly difficult to explain within the natural sciences. A book, for example, could be described physically in terms of the arrangements of its molecules, the kinds of atoms that it comprises, its chemical compounds. We could describe its mass and volume, and measure it for electrical and magnetic properties. Yet, these descriptions produce no hint as to a book’s subject matter. Only when the patterns of ink are considered, not as patterns of ink, but as *words*, does it become clear what a book is about.

Few, if any, things in the natural world have this property of aboutness. It makes no sense to ask what a stone is about, or what a river is about. While it makes sense to ask what a book or a newspaper is about, it makes no sense to ask what its components, the ink and paper, are about. It *does* make sense to ask what mental or cognitive processes are about – we often say to one another ‘what are you thinking about?’ One way of expressing the aboutness of mental processes is to say that they involve **representations** – our thoughts *represent* possible states of affairs, our perceptions *represent* our immediate environment (generally, though not always, accurately).

The representational quality of mental processes was described by the philosopher of psychology Franz Brentano (1838–1917). Brentano believed that mental states comprise mental *acts* and mental *contents*. So, for example, my believing that Rosie, my pet cat, is lazy is a mental state – I am in the state of believing that Rosie is lazy. For Brentano, the state has a dual character: it comprises an act, corresponding to the act of believing, and a content, namely the content that Rosie is lazy. Brentano thought that mental states can differ, even if they involve the same mental act. So, for example, my believing that Rosie is lazy, and my believing that all cats are lazy, would represent two different mental states. The same act is common to both, but the beliefs are differentiated by their content: one is *about* Rosie, the other is *about* all cats.

The consequence for Brentano was that psychology needs to consider not only the internal features of the mind or brain, but also what these features are about or represent in the world. Perhaps now you can see why it is not straightforward to decide what kind of science cognitive psychology is. Whereas physics and chemistry study the material world of atoms and molecules (which do not have this representational quality), cognitive psychology studies mental states whose representational nature cannot be ignored. Consequently, cognitive psychology studies something intrinsically relational – something that spans what is in the mind and what it relates to in the world. Indeed, the issue of representation tends to distinguish the social sciences (such as sociology) from the natural sciences (like

physics). Cognitive psychology, focusing on both what is represented (the world) and what does the representing (the mind), does not fall neatly into either category.

## 5.2 Computation

In Section 3 we considered some of the technological and theoretical antecedents to cognitive psychology. What emerged from the advances concerning theories of information and computation was the view that computers process information, and provide a means for modelling and understanding the mind. As David Marr put it, ‘If ... vision is really an information processing task, then I should be able to make my computer do it ...’ (Marr, 1982, p.4).

Marr’s statement hints at a deep relation between the computer and the mind. If computers process information, and information processing is what characterizes minds, perhaps, at some deep level, the mind is computational. This claim provides a further key assumption of the cognitive approach: cognitive psychologists tend to view the mind as computational, as well as representational.

Von Eckardt (1993) suggests that there are two assumptions involved in construing the mind as computational. First, is a linking assumption – the assumption that the mind is a computational device of some kind, and that its capacities are computational capacities. The assumption serves to link minds (things which we wish to understand better) with computers (things which are already well understood). Second, is the system assumption: this fleshes out what is meant by a computational device. Generally, the assumption tends to be that computers are systems that represent information, input, store, manipulate and output representations, and operate according to rules. The two assumptions work together to provide a framework for understanding the (relatively) unknown mind in terms of the known computer.

Just as with the representational assumption, the assumption that minds are computational raises many questions. One of the more pressing for cognitive psychology has been the precise form that computational models should take. This is in fact a major debate within contemporary cognitive psychology, and the issue will be referred to in one way or another in many chapters in this book (especially in Chapters 16 and 17). Broadly speaking, there have been two main proposals as to the computational models we should use to understand the mind: symbolic models and connectionist models.

### 5.2.1 Symbol systems

One way of understanding the idea that the mind is both representational and computational has been to suggest that the mind is a symbol system. On this view the representational qualities of the mind are expressed via the claim that the mind is symbolic and contains symbols. So, for example, my mental state that Rosie is lazy might be described as involving symbols for Rosie and laziness. The symbols together represent what the belief is about. To say that the mind is computational is to say none other than the mind embodies (computational) mechanisms for manipulating these symbolic representations. My believing that Rosie is lazy would then involve my appropriately manipulating the symbol for Rosie and the symbol for laziness.

Newell and Simon (1976) were the first to propose that the mind is a symbol system. In their view, symbolic representations and their manipulation are the very building blocks of intelligent thought and action. Newell and Simon proposed many different properties of symbol systems, but we need consider only a few. Symbol systems should comprise a basic set of symbols that can be combined to form larger symbol structures (just as the symbols for ‘Rosie’ and ‘lazy’ could be combined to form the symbolic expression ‘Rosie is lazy’). Symbol systems should contain processes that operate on symbol structures to produce other symbol structures. Finally, symbol structures should represent, or be about, objects.

Newell and Simon’s proposal that the mind is a symbol system amounts to the claim that the cognitive processes that underlie language, perception, memory, thinking, categorization, and problem solving will ultimately turn out to involve processes of manipulating and transforming symbolic representations. The proposal is, of course, an empirical one, and in principle the evidence could turn out either way. One way of addressing the issue is to develop models of symbol systems and compare these with empirical data (e.g. from human participants in an experiment). As you will see throughout this book, the strategy of producing computer models and comparing their performance with human data is a common one (see especially Chapter 16 for such comparisons for symbolic models). However, it is worth noting that disagreement with empirical evidence does not necessarily imply that the cognitive processes in question are not symbolic. It may well be that a different symbolic model would agree with the data much better. So, although the claim that the mind is a symbol system is empirical, it will require a considerable amount of empirical evidence to show either that the mind is symbolic or that it is not.

### 5.2.2 Connectionism

Cognitive psychologists have also sought to understand the mind’s representational and computational qualities via an alternative framework, known as connectionism.

Connectionist models typically draw their inspiration from some of the known characteristics of the brain. So, for example, we know that neurons are highly interconnected. Seemingly they can pass information on to neurons with which they are connected, either through inhibiting or enhancing the activity of those neurons. They appear to be able to process information in parallel – neurons are capable of firing concurrently. And there are many more properties besides. Connectionism describes attempts to build models of cognition out of building blocks that preserve these important properties of neural information processing. Typically, researchers simulate connectionist networks on a computer, networks that involve a number of layers of neuron-like computing units. The appeal of connectionism lies in the hope that connectionist models may ultimately stand a better chance of being successful models of cognition.

Consider the process of constructing symbolic and connectionist models in the area of language understanding, for example. A symbolic modeller might first seek to understand the representations involved in understanding language. They might posit symbolic representations of words and their meaning, of rules of grammar, and so on. They would then construct a computer program to encode the representations and manipulate them so that the program behaves sensibly. Given an input of written language, for example, the program might generate a representation of its meaning.

This would be an exceptionally hard task but, were it to be successful, we could then compare the output of the program with the judgments of human language understanders to see if the program generated sensible answers.

In contrast, a connectionist modeller, though trying to represent the same kinds of information, would do this in a different way. They would seek to represent information in terms of neuron-like computing units and their interconnections. Rather than freely writing a computer program, they would seek to explain language understanding in terms of the kinds of information processing that the neuron-like units engage in. Thus connectionists seek to restrict themselves to models that have some *prima facie* plausibility in terms of what we know of the information-processing properties of the brain.

One of the exciting findings associated with connectionism has been that this brain-like information processing tends to produce interesting cognitive properties all on its own (some properties do not have to be explicitly programmed, unlike the case of symbolic models). For example, people tend to be good at generalizing from just a few instances – though in all likelihood you have encountered few UK Prime Ministers, if you were asked to describe the typical UK Prime Minister you could probably come up with a sensible generalization (e.g. ambitious, driven, etc.). It turns out that connectionist models tend to be able to generalize quite spontaneously, with no need for this cognitive property to be explicitly programmed.

This brief discussion aimed only to introduce these different kinds of computational model; it has of course skated over many complexities. In particular, the question as to whether the mind is better modelled as a symbol system or as a connectionist network has been and continues to be hotly debated (see, for example, Fodor and Pylyshyn, 1988; Smolensky, 1987), as you will see especially in Chapters 16 and 17.

## Summary of Section 5

---

- Cognitive psychology is committed to the assumption that the mind is both representational and computational.
- Representations are understood as having a property of aboutness.
- Computations are understood as processes of inputting, storing, manipulating and outputting information.
- Within cognitive psychology, the mind tends to be understood in relation to either of two broad conceptions of computation:
  - computation as rule-based, symbol manipulation
  - computation as neurally-inspired, as in connectionist networks.

## 6 Level-dependent explanations

Linking the mind with computers raises many interesting and challenging questions. One view, commonly attributed to Marr (1982), is that cognition can be understood at, at least, three different levels.

### 6.1 The computational level

The first of Marr's level's (level 1) is commonly referred to as the computational level. An explanation of cognition at this level specifies *what* a computational system actually computes and *why*. The specification can be given in terms of a mapping between appropriate sets of inputs and their corresponding outputs. Consider a system that performs addition. A level 1 explanation would therefore refer to the 'plus' function, partially indicated in Table 1.2.

**Table 1.2** Level 1 specification for addition. The inputs are pairs of numbers to be added and the output is their sum

Inputs	Outputs
0,0	0
0,1	1
1,0	1
2,3	5
87,123	210

Marr also believed that level 1 explanations should specify why the system should compute the function that it does in order to solve a particular task. Why it is, for example, that the plus function (as opposed to multiplication) is the right function for the task of adding two numbers together?

Thus, cognitive psychologists that seek to explain some aspect of cognition at the computational level need to explain or describe the function that is computed (what the inputs and outputs are) and why that function is the appropriate one. For example, an explanation of language understanding might describe inputs that correspond to sentences or questions, and outputs that correspond to appropriate comments or responses.

### 6.2 The algorithmic level

Marr's level 2, commonly referred to as the algorithmic level, specifies *how* a computation is to be achieved. A level 2 explanation might describe the representations of input and output that a system employs, and the algorithms that operate over these representations. For example, in computing the 'plus' function, input numbers could be represented in a number of different ways: in denary or binary notation, as arabic or roman numerals, or as appropriate numbers of dots. The algorithm specifies the steps involved in transforming the input representations into appropriate output representations.

To return to the example of addition, one way of representing two numbers (say, the numbers 2 and 3) involves representing them in terms of appropriate numbers of dots (i.e. ●● and ●●●). One algorithm for adding the numbers might involve moving

the two dots one at a time so that they are adjacent to the three, to yield an output representation (not dissimilar to adding using an abacus). Another (formally) distinct algorithm would be to move the three dots one at a time so that they are adjacent to the two. These algorithms, and the sequence of steps they would generate, are shown in Table 1.3.

**Table 1.3** Two algorithms and the steps they generate for computing  $2 + 3$

<b>Algorithm 1 (move one dot at a time from right to left)</b>		<b>Step</b>	<b>Algorithm 2 (move one dot at a time from left to right)</b>	
Left	Right		Left	Right
••	•••	0	••	•••
•••	••	1	•	••••
••••	•	2		•••••
•••••		3		

Note that these two sequences of steps achieve the same end result, 5 dots (•••••) representing the number 5. That is, though they are distinct processes, and hence distinct algorithms, at level 1 they are indistinguishable. In fact, it can be proved that there are an infinite number of different algorithms for any level 1 specification.

This obviously makes it very difficult for a cognitive psychologist to work out what algorithm to choose in order to model human performance successfully. However, there are ways of distinguishing different algorithms. For example, algorithms can bestow a considerable benefit to anyone (or anything) that deploys them: even though a task may appear to be insoluble, or its solution appear to impose impractical demands on resources, with appropriate algorithms it may be soluble with a modicum of resources. Note how algorithm 2 in Table 1.3 completes the task in one less step than algorithm 1.

Less trivially, consider chess. One way of playing chess would be to consider all possible moves by looking ahead a certain number of steps. As one looks further ahead, however, the number of possible moves grows exponentially, and so this particular strategy would require vast amounts of memory and time. By deploying more sophisticated algorithms, ones involving heuristics and strategies that restrict the number of possible moves that need to be considered, the resource demands of the task fall rapidly. Thus, appropriate algorithms may render soluble tasks that appear insoluble, and also render them soluble within practical resource limits (Chapter 10 considers some of the stratagems of real chess experts).

To see this, consider different algorithms for multiplying 253 by 375. One option is to add 253 to itself 375 times. Another would involve adding 375 to itself 253 times. Yet another way would be to remember the products of all pairs of numbers up to, say, 400. The first and second algorithms would require a pencil and paper and a very large amount of time. By contrast, the third strategy would potentially require little time but a very large and efficient memory. A better algorithm, perhaps, would involve knowing by rote some products (say,  $5 \times 200 = 1,000$ ,  $3 \times 5 = 15$ , etc.), and knowing that the product asked for can be decomposed as follows:

$$\begin{aligned}
 253 \times 375 &= (200 + 50 + 3) \times (300 + 70 + 5) \\
 &= 200 \times (300 + 70 + 5) + 50 \times (300 + 70 + 5) + 3 \times (300 + 70 + 5) \\
 &= (60,000 + 14,000 + 1,000) + (15,000 + 3,500 + 250) + (900 + 210 + 15) \\
 &= 75,000 + 18,750 + 1,125 \\
 &= 94,875
 \end{aligned}$$

Note that this algorithm involves *some* demands on memory and *some* demands on time, but doesn't place excessive demands on either.

Returning to our example of language understanding, a challenge for a cognitive psychologist would be to work out how the inputs and outputs should be represented, and algorithms for converting the former into the latter. A critical question, however, will remain: why were these particular representations and this particular algorithm chosen, and could better choices have been made?

### 6.3 The implementational level

Marr's level 3 is commonly referred to as the hardware or implementational level. It specifies how algorithms and representations are physically realized. In our example of addition, numbers were realized as marks on pieces of paper and movement of those marks. In a digital computer, an explanation at the implementational level would make reference to transistors, voltages, currents, diodes and the like. If addition were implemented using an abacus, an explanation would make reference to beads sliding on rods. Were we to explain human cognitive processing in terms of Marr's level 3, then we would make reference to neurons, neurotransmitters and the like.

Explaining cognitive processing at the implementational level presents a very real challenge. In our example of language understanding, we would have to make reference to the real neural circuits that implement language understanding, and to their actual activities whilst doing so. Though neuropsychological and neuro-imaging evidence, as well as neuroscientific advances, accumulate, such an explanation exceeds the abilities of our current understanding.

### 6.4 Using Marr's levels

Cognitive psychologists tend to explain cognition at levels 1 and 2. That is, they pursue **functional accounts** (at level 1) and **process accounts** (at level 2). Level 3 explanations, those that refer to actual neurons, neurotransmitters and so on, tend to be left to neuroscientists. However, there are important relations between all three levels. For example, the implementation level can constrain what counts as an appropriate algorithm. The brain may not be able to implement all algorithms, or may not implement them equally well. In a sense, connectionist models are predicated on this view – that the hardware of the brain constrains our choice of algorithm (or level 2 explanations) to those that we know the brain is good at computing. Certainly if it could be shown that a level 1 or 2 account of some cognitive phenomenon could not be implemented in neural hardware, then real doubt would be cast on the corresponding psychological explanation.

This section has focused on some of the foundational assumptions made in contemporary cognitive psychology, though very many other assumptions are also made, and also tend to characterize a cognitive approach. Table 1.1 in Section 2 listed some of the more common ones, and you may wish to revisit it now. You may also like to refer to this table after you have read each of the following chapters to see if you can identify which assumptions have been made, and how explicitly.

## Summary of Section 6

---

- Marr's levels provide a framework for understanding explanations of cognition.
- Explanations can be pitched at one of three levels:
  - computational level
  - algorithmic level
  - implementational level.
- Cognitive psychological explanations are typically expressed at levels 1 (functional) and 2 (process), but are assumed to be constrained by what is known about level 3.

## 7 Conclusions

In the previous sections we have attempted to outline some of the history of cognitive psychology, its subject matter, and also some of its core assumptions. As we have seen, cognitive psychology has a relatively long history, and has made and continues to make many connections with other disciplines. To understand the nature of cognitive psychology, we have had to consider a wide range of issues, from computation to neuroimaging, from mundane but complex behaviour such as understanding language to the behaviour involved in anti-aircraft gunnery. Our survey has touched on action, perception, thinking, language, problem solving, categorization, and consciousness. We have considered the nature of scientific investigation, the importance of observation, and the need for, and practice of, sciences to posit theoretical entities that cannot be observed. We have also touched on the possibility that cognitive psychology may be a special science, perhaps somewhere between a social and a natural science.

### ACTIVITY 1.4

---

In Activity 1.1, we asked you to write down what you took to be the characteristic features of a scientific study of the mind. Take a few minutes to review your list – are there some features you would want to add to the list? And are there any you would want to remove?

---

In such a short chapter we have omitted much, and this chapter should be regarded as a partial survey of the foundations of cognitive psychology, intended to help you

make the most of the chapters that follow. Most notably, we have barely touched on the different methods of cognitive psychology, though the following chapters make clear just how central these methods are to the cognitive approach.

We have not intended to suggest that cognitive psychology faces no real challenges or problems. Far from it. Most if not all of the topics we will consider in this book are still not fully understood – though cognitive psychology has proved remarkably successful so far, it remains to be seen just how well it will deliver such a full understanding. Indeed, while in topics such as attention and perception cognitive psychologists have made great progress, others, such as consciousness and emotion, still present real challenges. This is not to say that cognitive psychologists have not contributed greatly. Indeed, as you will see in Chapters 13, 14 and 15 among others, progress has been made even though foundational questions remain.

The breadth of the many issues we have raised, as well as the results and promise of the cognitive approach that you will encounter in subsequent chapters, testify to the importance of developing a systematic and rigorous understanding of the mind. It also hints at the fascination and enjoyment that can be gained from studying cognitive psychology, something that we hope you will soon experience for yourself.

## Further reading

- Bechtel, W. (1988) *Philosophy of Mind: An Overview for Cognitive Science*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Bechtel W. and Abrahamsen A. (1991) *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Oxford, Blackwell.
- Gardner, H. (1985) *The Mind's New Science: A History of the Cognitive Revolution*, New York, Basic Books.

## References

- Bryan, W.L. and Harter, N. (1899) ‘Studies on the telegraphic language: the acquisition of a hierarchy of habits’, *Psychological Review*, vol.6, pp.345–75.
- Chomsky, N. (1959) ‘A review of B.F. Skinner’s “Verbal Behavior”’, *Language*, vol.35, no.1, pp.26–58.
- Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M. and Tyler, L.K. (2000) ‘Susceptibility induced loss of signal: comparing PET and fMRI on a semantic task’, *NeuroImage*, vol.11, pp.589–600.
- Fodor, J.A. (1974) ‘Special sciences’, *Synthese*, vol.28, pp.77–115.
- Fodor, J.A. and Pylyshyn, Z.W. (1988) ‘Connectionism and cognitive architecture: a critical analysis’, *Cognition*, vol.28, pp.3–71.
- Greer, M.J., van Casteren, M., McLellan, S.A., Moss, H.E., Rodd, J., Rogers, T.T. and Tyler, L.K. (2001) ‘The emergence of semantic categories from distributed featural representations’, *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, London, Lawrence Erlbaum Associates.

- Hillis, A.E. and Caramazza, A. (1991) ‘Category-specific naming and comprehension impairment: a double dissociation’, *Brain Language*, vol.114, pp.2081–94.
- Lashley, K.S. (1951) ‘The problem of serial order in behaviour’, in Jeffress, L.A. (ed.) *Cerebral Mechanisms in Behaviour: The Hixon Symposium*, New York, John Wiley.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman & Company.
- Miller, G.A., Galanter, E. and Pribam, K. (1960) *Plans and the Structure of Behaviour*, New York, Holt, Rinehart and Winston.
- Newell A. and Simon H.A. (1976) ‘Computer science as empirical enquiry: symbols and search’, *Communications of the Association for Computing Machinery*, vol.19, pp.113–26.
- Skinner, B.F. (1957) *Verbal Behavior*, New York, Appleton-Century-Crofts.
- Smolensky P. (1987) ‘The constituent structure of connectionist mental states: a reply to Fodor and Pylyshyn’, *The Southern Journal of Philosophy, Supplement*, vol.26, pp.137–61.
- Tolman, E.C. (1932) *Purposive Behaviour in Animals and Man*, New York, Century.
- Turing A.M. (1950) ‘Computing machinery and intelligence’, *Mind*, vol.LIX, pp.433–60.
- Tyler, L.K. and Moss, H.E. (2001) ‘Towards a distributed account of conceptual knowledge’, *Trends in Cognitive Sciences*, vol.5, no.6, pp.244–52.
- Warrington, E.K. and Shallice, T. (1984) ‘Category specific semantic impairments’, *Brain*, vol.107, pp.829–54.
- Wertheimer, M. (1912) ‘Experimentelle Studien über das Sehen von Bewegung’, *Zeitschrift Für Psychologie*, vol.61, pp.161–265.
- Von Eckardt, B. (1993) *What Is Cognitive Science?*, Cambridge, MA, MIT Press.

---

# PART 1

# PERCEPTUAL PROCESSES

---

Introduction

**Chapter 2 Attention**

*Peter Naish*

**Chapter 3 Perception**

*Graham Pike and Graham Edgar*

**Chapter 4 Recognition**

*Graham Pike and Nicola Brace*

# Introduction

In Part 1 you will find chapters on attention, perception and recognition. Why do we begin with these particular topics? Well, there is a fairly strong tradition of placing these topics early in books on cognition, and there are at least two reasons for this. First, there is a strong applied psychology theme to all these topics, whether it is finding better ways to present relevant information to people in safety-critical occupations, such as aircraft pilots, devising techniques for improving eye-witness identification, or designing machines that can ‘see’ and ‘recognize’. Second, attention, perception and recognition are all topics that concern the relationship between the mind and the world, which seems a good place to start trying to understand the mind itself. Other chapters – for example, Chapter 6 on language processing – also address the issue of how information from the world gets ‘into’ the mind, but the topics of attention, perception and recognition provide particularly direct questions relating to it. Why do we become aware of some aspects of the environment rather than others? How is it that we manage to perceive those things we do become aware of? And for those things we do consciously perceive, how do we come to recognize what they are?

As you will see, these turn out to be far from simple questions and to require far from simple answers. A key issue that comes up in all three chapters has to do with distinguishing between aspects of the world (physics), how these aspects affect the body and especially the nervous system (physiology), and what mental representations result (cognitive psychology). In Chapter 2, you will learn what kinds of physical energy the auditory system uses to represent the location of a sound source; in Chapter 3 you will encounter a theory of how the visual system comes to represent gestalt organization, which is easily mistaken for a property of the world rather than of the mind; in Chapter 4 you will see how different aspects of the same physical face – familiarity, identity, emotional tone – are processed by different physiological pathways and have separate cognitive representations.

A further key issue that emerges in all three chapters is the fractionation of functions. It turns out that there is not just one sort of attention but many different forms of it. Similarly, it transpires that visual perception is far from being a unitary function; in fact, vision is made up of such a multitude of component processing streams that Chapter 3 has space to mention only some of them. As indicated in the previous paragraph, recognition can also be analysed into different processes, and a similar fractionation will recur in later chapters in relation to other mental functions such as memory. (How we should conceptualize all these cognitive functions and their sub-components is something it might be useful to consider in the light of Chapter 5 on categorization.) Allied to the issue of how cognitive functions can be analysed into component processes are questions as to which of these processes result in representations that are or are not consciously experienced, and which can be carried out in parallel and which only one at a time.

A common theme across all the chapters is the use of neuropsychological evidence to help elucidate key issues such as those we have just identified. Injury to the brain can affect attention, perception and recognition in quite unexpected ways. Studying the behavioural and phenomenological consequences of injury to specific parts of the brain, relating neuroanatomy to behaviour and conscious experience,

throws light upon the structure of cognition by providing both tests of psychological theories and grounds from which theories may be derived.

Another issue common to all the chapters is the extent to which stored knowledge enters into the functions of attention, perception and recognition. These functions might be purely stimulus-driven; that is, driven by physical properties of the world. But if they are not, then at what stage in processing does prior knowledge exert its influence? Do we, for example, necessarily identify a plant *before* picking it? If not, why would we tend to avoid picking stinging nettles with bare hands? Do we perceive familiar faces in the same way that we perceive unfamiliar faces? If not, does familiarity also affect perception of other classes of object? It is important that answers to such questions are given within a theoretical context. When you have read the chapters, you should reflect on how well or how badly cognitive psychological theories have fared in recent decades.

In Chapter 2, Peter Naish describes such different forms of attention as attention to regions of space, attention to objects and attention for action, but attempts finally to summarize them all under a single fairly abstract definition of the term. He shows how ideas about attention have changed and diversified over the last fifty years and considers how well the early theories have stood up to examination. In Chapter 3, Graham Pike and Graham Edgar consider top-down and bottom-up theories of perception, and propose a resolution in terms of perception for recognition and perception for action. They also introduce and evaluate Marr's computational framework for a bottom-up theory of perception. Lastly, in Chapter 4, Graham Pike and Nicola Brace describe and contrast two theories of object perception, as well as a model of face perception that has been implemented as a connectionist network. Across the chapters you will encounter theories being tested and sometimes confirmed and sometimes found wanting. You will also meet the idea that different theories may be complements of one another rather than simply alternatives. One theory may succeed in one domain but fail in another, and vice versa for a second theory. You will also see how confidence in a theory varies with its range of application, and how confidence can be boosted if it proves possible to implement the theory as a working computer model. The challenge for the future is for theorists to develop more detailed and implementable theories of attention, perception and recognition whilst allowing that different people may find distinct ways of doing the same thing.



Peter Naish

---

### 1 Auditory attention

For many of us the concept of attention may have rather negative connotations. At school we were told to pay attention, making us all too aware that it was not possible to listen to the teacher while at the same time being lost in more interesting thoughts. Neither does it seem possible to listen effectively to two different things at the same time. How many parents with young children would love to be able to do that! One could be excused for feeling that evolution has let us down by failing to enable us to process more than one thing at a time. If that is how you feel, then this chapter might add insult to injury, because it will cite evidence that we do in fact process a good deal of the material to which we are not attending. Why, you might ask, do we go to the trouble of analysing incoming information, only to remain ignorant of the results? To attempt an answer it is necessary to consider a range of issues, stretching from registration of information by the sense organs, through the processes of perception, to the nature of awareness and consciousness. Attention is a broad and intriguing topic. That breadth makes it very difficult to offer a simple definition of the term, so I will not attempt to do so until the end of the chapter.

To cover some of this topic (we have only a chapter, and there are whole books on the subject) I shall follow an approximately historical sequence, showing how generations of psychologists have tackled the issues and gradually refined and developed their theories. You will discover that initially there seemed to them to be only one role for attention, but that gradually it has been implicated in an ever-widening range of mental processes. As we work through the subject, two basic issues will emerge. One is concerned with the mechanisms of attention, and raises questions such as:

- How much material can we take in at once?
- What happens to information to which we did not attend?
- In what circumstances does attention fail, allowing unwanted information to influence or distract us?

The other theme has a more philosophical flavour, and raises questions concerning why we experience the apparent limitations of attention:

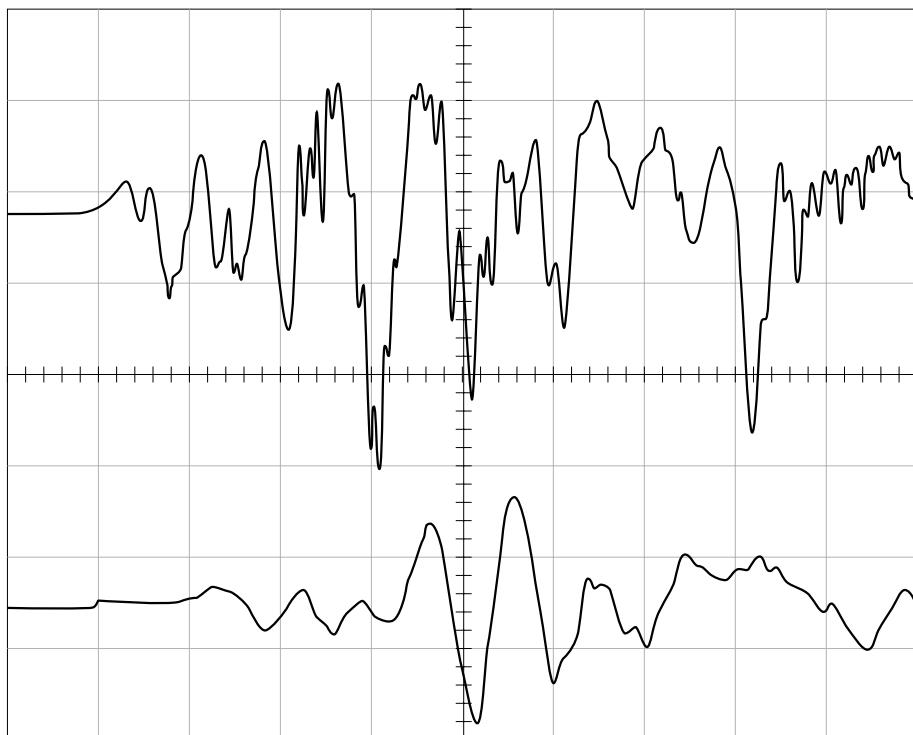
- Are the limitations simply an inevitable characteristic of a finite brain?
- Have we evolved to exhibit attention – that is, does it confer advantages?

We shall begin to explore these issues by looking at the ways in which one of our senses (hearing) has developed to facilitate attention.

#### 1.1 Disentangling sounds

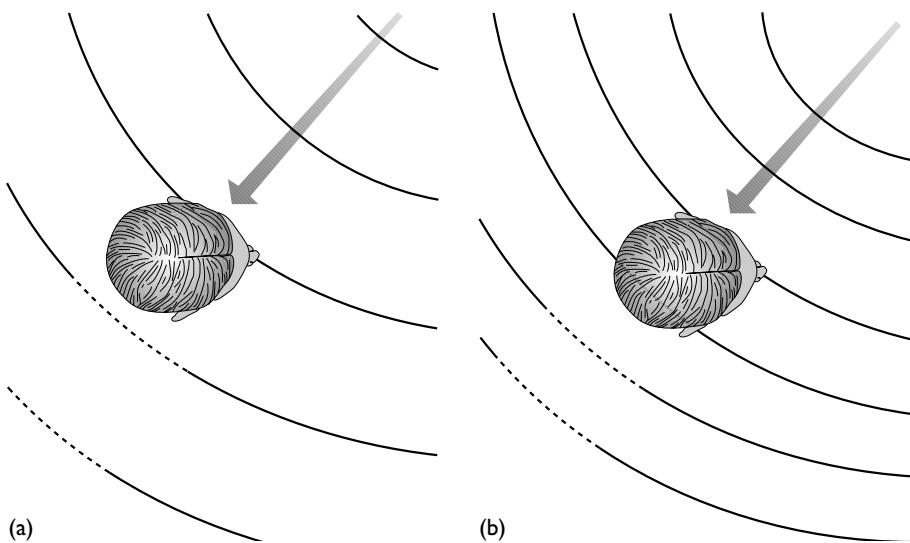
If you are still feeling aggrieved about the shortcomings of evolution, then you might take heart from the remarkable way in which the auditory system has evolved so as to

avoid a serious potential problem. Unlike our eyes, our ears cannot be directed so as to avoid registering material that we wish to ignore; whatever sounds are present in the environment, we must inevitably be exposed to them. In a busy setting such as a party we are swamped by simultaneous sounds – people in different parts of the room all talking at the same time. An analogous situation for the visual system would be if several people wrote superimposed messages on the same piece of paper, and we then attempted to pick out one of the messages and read it. Because that kind of visual superimposition does not normally occur, there have been no evolutionary pressures for the visual system to find a solution to the problem (though see below). The situation is different with hearing, but the possession of two ears has provided the basis for a solution.



**Figure 2.1** The waveform of a hand clap, recorded at the left (upper trace) and right (lower trace) ears. Horizontal squares represent durations of 500 microseconds (a microsecond is one-millionth of a second); vertical divisions are an arbitrary measure of sound intensity

Figure 2.1 shows a plot of sound waves recorded from inside a listener's ears. You can think of the up and down movements of the wavy lines as representing the in and out vibrations of the listener's ear drums. The sound was of a single hand clap, taking place to the front left of the listener. You will notice that the wave for the right ear (i.e. the one further from the sound) comes slightly later than the left (shown by the plot being shifted to the right). This right-ear plot also goes up and down far less, indicating that it was less intense, or in hearing terms that it sounded less loud at that ear. These differences, in timing and intensity, are important to the auditory system, as will be explained.



**Figure 2.2** Curved lines represent wave crests of a sound approaching from a listener's front left. In (b) the sound has a shorter wavelength (higher pitched) than in (a), so waves are closer together, with a crest at each ear

Figure 2.2(a) represents sound waves spreading out from a source and passing a listener's head. Sound waves spread through the air in a very similar way to the waves (ripples) spreading across a pond when a stone is thrown in. For ease of drawing, the figure just indicates a 'snapshot' of the positions of the wave crests at a particular moment in time. Two effects are shown. First, the ear further from the sound is slightly shadowed by the head, so receives a somewhat quieter sound (as in Figure 2.1). The head is not a very large obstacle, so the intensity difference between the ears is not great; however, the difference is sufficient for the auditory system to register and use it. If the sound source were straight ahead there would be no difference, so the size of the disparity gives an indication of the sound direction. The figure also shows a second difference between the ears: a different wave part (crest) has reached the nearer left ear than the further right ear (which is positioned somewhere in a trough between two peaks). Once again, the inter-aural difference is eliminated for sounds coming from straight ahead, so the size of this difference also indicates direction.

Why should we make use of both intensity and wave-position differences? The reason is that neither alone is effective for all sounds. I mentioned that the head is not a very large obstacle; what really counts is how large it is compared with a **wavelength**. The wavelength is the distance from one wave crest to the next. Sounds which we perceive as low pitched have long wavelengths – longer in fact than the width of the head. As a result, the waves pass by almost as if the head was not there. This means that there is negligible intensity shadowing, so the intensity cue is not available for direction judgement with low-pitched sounds. In contrast, sounds which we experience as high pitched (e.g. the jingling of coins) have wavelengths that are shorter than head width. For these waves the head is a significant obstacle, and shadowing results. To summarize, intensity cues are available only for sounds of short wavelength.

In contrast to the shadowing effect, detecting that the two ears are at different positions on the wave works well for long wavelength sounds. However, it produces ambiguities for shorter waves. The reason is that if the wave crests were closer than the distance from ear to ear, the system would not be able to judge whether additional waves should be allowed for. Figure 2.2(b) shows an extreme example of the problem. The two ears are actually detecting identical parts of the wave, a situation which is normally interpreted as indicating sound coming from the front. As can be seen, this wave actually comes from the side. Our auditory system has evolved so that this inter-ear comparison is made only for waves that are longer than the head width, so the possibility of the above error occurring is eliminated. Consequently, this method of direction finding is effective only for sounds with long wavelengths, such as deeper speech sounds.

You will notice that the two locating processes complement each other perfectly, with the change from one to the other taking place where wavelengths match head width. Naturally occurring sounds usually contain a whole range of wavelengths, so both direction-sensing systems come into play and we are quite good at judging where a sound is coming from. However, if the only wavelengths present are about head size, then neither process is fully effective and we become poor at sensing the direction. Interestingly, animals have evolved to exploit this weakness. For example, pheasant chicks (that live on the ground and cannot fly to escape predators) emit chirps that are in the ‘difficult’ wavelength range for the auditory system of a fox. The chicks’ mother, with her bird-sized head, does not have any problems at the chirp wavelength, so can find her offspring easily. For some strange reason, mobile telephone manufacturers seem to have followed the same principle. To my ears they have adopted ringtones with frequencies that make it impossible to know whether it is one’s own or someone else’s phone which is ringing!

## ACTIVITY 2.1

---

- 1 Set up a sound source (the radio, say), then listen to it from across the room. Turn sideways-on, so that one ear faces the source. Now place a finger in that nearer ear, so that you can hear the sound only via the more distant ear. You should find that the sound seems more muffled and deeper, as if someone had turned down the treble on the tone control. This occurs because the shorter wavelength (higher pitched) sounds cannot get round your head to the uncovered ear. In fact you may still hear a little of those sounds, because they can reflect from the walls, and so reach your uncovered ear ‘the long way round’. Most rooms have sufficient furnishings (carpets, curtains, etc.) to reduce these reflections, so you probably will not hear much of the higher sounds. However, if you are able to find a rather bare room (bathrooms often have hard, shiny surfaces) you can use it to experience the next effect.
- 2 Do the same as before, but this time you do not need to be sideways to the sound. If you compare your experiences with and without the finger in one ear you will probably notice that, when you have the obstruction, the sound is more ‘boomy’ and unclear. This lack of clarity results from the main sound, which comes directly from the source, being partly smothered by slightly

later echoes, which take longer routes to your ear via many different paths involving reflections off the walls etc. These echoes are still there when both ears are uncovered, but with two ears your auditory system is able to detect that the echoes are coming from different directions from the main sound source, enabling you to ignore them. People with hearing impairment are sometimes unable to use inter-aural differences, so find noisy or echoing surroundings difficult.

---

## 1.2 Attending to sounds

From the above, you will appreciate that the auditory system is able to separate different, superimposed sounds on the basis of their different source directions. This makes it possible to attend to any one sound without confusion, and we have the sensation of moving our ‘listening attention’ to focus on the desired sound. For example, as I write this I can listen to the quiet hum of the computer in front of me, or swing my attention to the bird song outside the window to my right. Making that change feels almost like swinging my eyes from the computer to the window and the term **spotlight of attention** has been used to describe the way in which we can bring our attention to bear on a desired part of the environment.

My account so far has explained the mechanisms that stop sounds becoming ‘jumbled’ and reminds us that, subjectively, we listen to just one of the disentangled sounds. It seems obvious that they would need disentangling to become intelligible, but why do we then attend to only one? That question leads us into the early history of attention research.

One of the first modern researchers formally to investigate the nature of auditory attention was Broadbent (1952, 1954), who used an experimental technique known as **dichotic listening**. This offers a way of presenting listeners with a simplified, more easily manipulated version of the real world of multiple sounds. Participants wear a pair of headphones, and receive a different sound in each ear; in many studies the sounds are recorded speech, each ear receiving a different message. Broadbent and others (e.g. Treisman, 1960) showed that, after attending to the message in one ear, a participant could remember virtually nothing of the unattended message that had been played to the other, often not even the language spoken.

Broadbent’s experiments showed that two refinements should be made to the last statement. First, if the two messages were very short, say just three words in each ear, then the participant could report what had been heard by the unattended ear. The system behaved as if there were a short-lived store that could hold a small segment of the unattended material until analysis of the attended words was complete. Second, if the attended message lasted more than a few seconds, then the as yet unprocessed material in the other ear would be lost. The store’s quality of hanging on to a sound for a short time, like a dying echo, led to it being termed the **echoic memory**.

It was also shown that people would often be aware of whether an unattended voice had been male or female, and they could use that distinction to follow a message. Two sequences of words were recorded, one set by a woman, the other by a man. Instead of playing one of these voice sequences to each headphone, the words were made to alternate. Thus, the man’s voice jumped back and forth, left to right to

left, while the woman's switched right to left to right. In this situation participants were able to abandon the normal 'attending by ear' procedure, and instead report what a particular speaker had said; instead of using location as a cue for attention, they were using the pitch of the voice.

The explanation for these findings seemed straightforward. Clearly the brain had to process the information in a sound in order to understand it as speech. In this respect, the brain was rather like a computer processing information (computers were beginning to appear at that time), and everyone knew that computers could only process one thing at a time – that is, **serially**. Obviously (theorists thought) the brain must be serial too, so, while processing the information of interest, it needed to be protected from all the rest: it needed to attend and select. However, the earliest stages of processing would have to take place in **parallel** (i.e. taking in everything simultaneously), ensuring that all information would potentially be available, but these initial processes would have to utilize very simple selection procedures; anything more complex would demand serial processing. The procedures were indeed simple: attention was directed either on the basis of the direction of a sound, or on whether it was higher or lower pitched. Broadbent's (1954) theory was that, after the first early stage of parallel information capture, a 'gate' was opened to one stream of information and closed to the rest.

## 2.1

### Research study

#### Application of research on auditory attention

Donald Broadbent's early career included research for the UK Ministry of Defence, and his findings often led to innovation. One problem he addressed was the difficulty pilots experienced, when trying to pick out a radio message from a number of interfering stations (radio was less sophisticated then). Pilots' headphones delivered the same signals to each ear, so it was not possible to use inter-aural differences to direct attention to the wanted message. Broadbent devised a stereo system, which played the desired signal through *both* headphones, while the interference went only to one or the other. This made the interference seem to come from the sides, while the signal sounded as if it was in the middle (identical waves at the two ears). In effect, this was dichotic listening, with a third (wanted) signal between the other two. The improvement in intelligibility was dramatic, but when Broadbent played a recording to officials they decided that it was so good that he must have 'doctored' the signal! The system was not adopted. Decades later, I demonstrated (Naish, 1990) that using stereo, and giving a directional quality to the headphone warning sounds used in aircraft cockpits, could result in significantly shorter response times. Thus, the warning indicating an approaching missile could be made to seem as if coming from the missile direction, so speeding the pilot's evasive measures. The next generation of fighter aircraft may at last incorporate '3-D' sound.

### 1.3 Eavesdropping on the unattended message

It was not long before researchers devised more complex ways of testing Broadbent's theory of attention, and it soon became clear that it could not be entirely correct. Even in the absence of formal experiments, common experiences might lead one to question the theory. An oft-cited example is the **cocktail party effect**. Imagine you are attending a noisy party, but your auditory location system is working wonderfully, enabling you to focus upon one particular conversation. Suddenly, from elsewhere in the room, you hear someone mention your name! If you were previously selecting the first conversation, on the basis of its direction and the speaker's voice, then how did your 'serial' brain manage to process another set of sounds in order to recognize your name?

Addressing this puzzle, Treisman (1960) suggested that, rather than the all-or-nothing selection process implied by Broadbent, the ability to pick out one's name could be explained by an **attenuation** process. The attenuation process would function as if there were a **filter**, 'turning the volume down' for all but the attended signal. Although that would leave most unattended material so attenuated as to be unnoticed, for a signal to which we were very sensitive, such as our own name, there would be sufficient residual information for it to be processed and hence attract our attention. Treisman devised a series of ingenious experiments which supported this idea. Many of her studies involved **shadowing**, a dichotic listening technique which requires the participant to repeat aloud everything that is heard in one ear, following like a shadow close behind the spoken message. (NB this is not to be confused with the very different 'head shadowing' referred to earlier.) This task demands concentration, and when the shadowed message ceases the participant appears to be completely ignorant of what was said in the other ear.

In one experiment Treisman actually made the storylines in the messages swap ears in the middle of what was being said. Thus, the left ear might hear:

*Little Red Riding Hood finally reached the cottage, but the wicked wolf was in \* beds; one was large, one medium and one small.*

Meanwhile, the right ear would receive:

*When she had finished the porridge, Goldilocks went upstairs and found three \* bed, dressed in the grandmother's clothes.*

The asterisks indicate where the storylines swap ears. The interesting finding is that when asked to shadow one ear participants tend to end by shadowing the other, because they follow the sense of the story. Broadbent's position could not explain that, since the listener could not know that the story continued in the other ear, if that ear had been completely ignored. Treisman, on the other hand, claimed that the story temporarily sensitized the listener to the next expected words, just as with the permanent sensitization associated with our own name. Sensitization of this temporary kind is known as **priming**, and many experimental techniques have demonstrated its existence. For example, in a **lexical decision task** (a task that requires participants to indicate as quickly as possible whether or not a string of letters spells a real word), people can respond much more quickly to a word if it is preceded by another related to it. For example, the 'Yes' is given to *doctor* (yes, because it is a word) more quickly when presented after the word *nurse* than when following the word *cook*.

Treisman's ideas stimulated a succession of experiments, some seeming to show that information could 'get through' from a wider range of stimuli than one's own name or a highly predictable word in a sentence. For example, Corteen and Wood (1972) carried out a two-part experiment. Initially they presented their participants with a series of words, and each time a word from a particular category (city name) appeared the participant was given a mild electric shock. In this way, an association was formed between the shock and the category. Although the shocks were not really painful, they inevitably resulted in something like mild apprehension when one of the critical words was presented. This response (which once learned did not require the shocks in order for it to continue) could be detected as a momentary change in skin electrical resistance. The sweat glands of a nervous person begin to secrete, and the salty fluid lowers the resistance to a small (non-shocking) electric current. The change is known as the **galvanic skin response (GSR)** and has been used in so-called lie detectors. Corteen and Wood connected their participants to GSR apparatus when they started the second part of the experiment: a dichotic listening task. As usual, participants could later remember nothing about the unattended message, but the GSR showed that each time the ignored ear received one of the 'shocked' words there was a response. Moreover, a GSR was detected even to words of the same category, but which had not been presented during the shock-association phase. This generalizing of the response to un-presented words strengthens the claim that their meanings were established, even when not consciously perceived.

Not surprisingly, at this stage of research into auditory attention a number of psychologists began to question the idea that the brain could not process more than one signal at a time. Deutsch and Deutsch (1963) suggested that *all* messages received the same processing, whether they were attended or not; Norman (1968) proposed that unattended information must at least receive sufficient processing to activate relevant semantic memories (i.e. the memory system that stores the meanings of words; see Chapter 8). These suggestions certainly explained the intriguing dichotic listening results, showing people to be influenced by material of which they seemed to have no knowledge. However, the ideas, if true, would require the brain to be far more parallel in its function than had been supposed. At that time there was neither an analogue by which parallel processing could be conceptualized, nor sufficient neuroanatomical information to contribute to the debate. Today there is ample evidence of the parallel nature of much of the brain's processing and, additionally, computers have advanced to the stage where brain-like parallel processing can be emulated (see Chapter 16). Thus, modern researchers have no difficulty in conceptualizing parallel processing and the nature of the attention debate has shifted somewhat. Nevertheless, recent studies have also revealed that early stages of analysis are modified by attention, effects that Broadbent would have immediately recognized as examples of filtering. We shall explore these issues in more depth, after first considering the nature of attention in visual processing.

## Summary of Section 1

---

The auditory system is able to process sounds in such a way that, although several may be present simultaneously, it is possible to focus upon the message of interest. However, in experiments on auditory attention, there have been contradictory results concerning the fate of the unattended material:

- The auditory system processes mixed sounds in such a way that it is possible to focus upon a single wanted message.
- Unattended material appears not to be processed:
  - The listener is normally unable to report significant details concerning the unattended information.
  - Only the most recent unattended material is available, while still preserved in the echoic memory.
- These results suggest parallel acquisition of all available information, followed by serial processing to determine meaning for one attended message.
- Although there is little conscious awareness of unattended material, it may receive more processing than the above results imply:
  - Words presented to the unattended ear can produce priming and physiological effects.
  - Participants trying to ‘shadow’ one ear will follow the message to the other ear.
- These results imply that processing takes place in parallel, to the extent that meaning is extracted even from unattended material.

## 2 Visual attention

I introduced Section 1 by suggesting that the auditory system had a special problem: unlike the visual system, it needed processes which would permit a listener to attend to a specific set of sounds without being confused by the overlap of other, irrelevant noises. The implication of that line of argument was that vision had no need of any such system. However, although we do not see simultaneously *everything* that surrounds us, we can certainly see more than one thing at a time. Earlier, I wrote of attending to the sound of the computer in front of me, or of the birds to one side. I can do much the same visually. While keeping my eyes directed to the computer screen, I can either attend to the text I am typing or, out of the corner of my eye, I can be aware of the window and detect a bird when it flies past. If our eyes can receive a wide range of information in parallel, does that give the brain an attentional problem analogous to that of disentangling sounds? If visual information is handled in much the same way as auditory information seems to be, then we might expect the various items in the field of view to activate representations in memory simultaneously. That should lead to effects equivalent to those found in listening experiments; in other words, it might be possible to show that we are influenced by items which we did not

know we had seen. We shall examine evidence of this shortly, but I shall first draw your attention to another area of similarity between hearing and seeing.

I pointed out at the start of Section 1.1 that, whereas we often have to follow one speech stream while ignoring others, we do not normally have to disentangle overlapping handwriting. However, it is worth bearing in mind that visual objects do overlap and hide parts of each other, and the brain certainly has the problem of establishing which components of the image on the retina ‘go together’ to form an object. This issue is examined in more depth in Chapter 3.

As with hearing, a variety of cues is available to help in directing visual attention. Taking my window again as an example, I can either look at the glass and see a smear (I really must get round to washing the window!), or I can look through that, to the magpie sitting chattering in the apple tree. In this kind of situation we use distance to help separate objects, in much the same way as we use direction in hearing. However, we can deploy our attention in a more sophisticated way than simply on the basis of distance, as can be demonstrated by another aircraft-related example.

Military jets are often flown very fast and close to the ground (to avoid radar detection), requiring the pilot to attend intently to the outside view. At the same time, there are various pieces of information, traditionally displayed on instruments within the cockpit, which the pilot must check frequently. To avoid the pilot having to look down into the cockpit, the ‘head-up display’ (HUD) was developed. This comprises a piece of glass, just in front of the pilot, in which all the vital information is reflected. The pilot can read the reflection, or look through it to the outside world, just as one can look at reflections in a shop window, or look through to the goods on display. With a simple reflection, the pilot would still have to change focus, like me looking at the smear or the bird. However, modern HUDs use an optical system which makes the information reflected in the display appear to be as far away as the outside scene. This saves valuable re-focusing time. Nevertheless, although the numerals in the HUD now appear to be located at the same distance as, say, a runway, pilots still have the sensation of focusing on one or the other; if they are reading their altitude they are relatively unaware of the scene on which it is superimposed. This suggests (as we shall see in more detail later) that visual attention can be linked to specific objects rather than to general regions of space, very much as auditory attention can follow a particular speaker’s voice, or the sense of a sentence.

## 2.1 Knowing about unseen information

An obvious difference between hearing and seeing is that the former is extended in time, while the latter extends over space. So, for example, we can listen to a spoken sentence coming from one place, but it takes some time to hear it all. In contrast, a written sentence is spread over an area (of paper, say) but, as long as it is reasonably short, it can be seen almost instantly. Nevertheless, seeing does require some finite time to capture and analyse the information. This process can be explored by presenting letters or words for a short, measured period of time; nowadays they are shown on a computer screen, but early research used a dedicated piece of apparatus, called a tachistoscope. Just how long was required to register a small amount of information was investigated by Sperling (1960), who showed participants grids of letters, arranged as three rows of four letters each. If such a display was presented for 50 ms (i.e. 50 milliseconds, which is one twentieth of a second), people were

typically able to report three or four of the letters; the rest seemed to have remained unregistered in that brief period of time.

Sperling explored this further. He cued participants with a tone, indicating which of the three rows of letters they should try to report; a high note for the top row, lower for middle and deep for bottom. Crucially, the tones were not presented until just *after* the display had disappeared, meaning that participants were not able to shift their attention in preparation for the relevant row of letters when presented: it already had been presented. Strange as it seemed, people were still able to report three or four items from the cued row. Since they did not know until after the display had gone which row would be cued, this result implied that they must have registered most of the letters in *every row*; in other words, between nine and 12 letters in total. This apparent paradox, of seeming to know about a larger proportion of the items when asked only to report on some of them, is called the **partial report superiority effect**. The effect was also observed if letters were printed six in red and six in black ink, then two tones used to indicate which colour to report. Participants seemed to know as much about one half (the red, say) as they did about all 12, implying that, although they could not report all the letters, there was a brief moment when they did have access to the full set and could choose where to direct their attention. The ‘brief moment’ was equivalent to the echoic memory associated with dichotic listening experiments, so the visual counterpart was termed an **iconic memory** (an icon being an image). All the material seemed to be captured in parallel, and for a short time was held in iconic memory. Some was selected for further, serial processing, on the basis of position or colour; these being analogous to position and voice pitch in dichotic listening tasks. Unselected material (the remaining letters) could not be remembered.

With the close parallels between these auditory and visual experiments, you will not be surprised to learn that the simple selection and serial processing story was again soon challenged, and in very similar ways. Where the hearing research used shadowing to prevent conscious processing of material, the visual experiments used **backward masking**. Masking is a procedure in which one stimulus (the target) is rendered undetectable by the presentation of another (the mask); in backward masking the mask is presented after the target, usually appearing in the order of 10–50 ms after the target first appeared. The time between the onset of the target display and the onset of the mask is called the **stimulus onset asynchrony** (SOA). The target might be an array of letters or words; this disappears after a few tens of milliseconds, to be replaced by the mask, which is often a random pattern of lines. The SOA can be adjusted until participants report that they do not even know whether there has been a target, let alone what it was. In such circumstances the influence of the masked material seems sometimes still to be detected via priming effects. Thus, Evett and Humphreys (1981) used stimulus sequences containing two words, both of which were masked. The first was supposed to be impossible to see, while the second was very difficult. It was found that when the second word was related to the first (e.g. ‘tiger’ following ‘lion’) it was more likely to be reported accurately; the first, ‘invisible’ word apparently acted as a prime.

Claims such as these have not gone unchallenged. For example, Cheesman and Merikle (1984) pointed out that although participants say they cannot see masked words, they often do better than chance when forced to guess whether or not one had actually been presented. These researchers insisted that proper conclusions about

extracting meaning from unseen material could be made only if the material was truly unseen; that is, when the participants could do no better than chance. Under these conditions they found no evidence for priming by masked words. However, more recently researchers have provided persuasive evidence that meaning *can* be extracted from material of which the participant is unaware. This is worth examining in more detail.

Pecher *et al.* (2002) used the Evett and Humphreys (1981) technique, but with modifications. As in the earlier study, they showed a potential prime (e.g. 'lion'), followed by a hard-to-see masked target (e.g. 'tiger'). However, there were two changes in this study. First, the priming word could be displayed either for a very short time, so that it was allegedly undetectable, or it was shown for a duration of 1 second, giving ample time for reading and guaranteeing a priming effect.

The second change was to use two sets of trials. In one, the following target was almost always (90 per cent of the time) related to the prime (e.g. 'lion' followed by 'tiger'). In the other set of trials only 10 per cent of trials used related words. For remaining trials the stimuli were unrelated, so that the first word was not strictly a prime (e.g. 'list' followed by 'tiger'). The results of this study are summarized in Table 2.1.

**Table 2.1** The percentage of targets correctly reported under various priming conditions

	<b>Short duration prime</b>		<b>1 second prime</b>	
	10% related	90% related	10% related	90% related
Related words	56	52	70	91
Unrelated words	49	43	55	51
Priming advantage	7	9	15	40

Source: adapted from Pecher *et al.*, 2002

The effects are best appreciated by looking first at the final two columns of figures, showing the results when the first word was displayed for 1 second. For the condition where only 10 per cent of targets were related to the preceding word, 70 per cent of those targets were correctly identified when there was a relationship. The hit rate fell to 55 per cent when the targets were not related, so the priming effect produced a 15 per cent advantage ( $70 - 55 = 15$ ). The last column shows a massive 91 per cent hit rate for related words, when there was a 90 per cent chance that they would be related to the preceding prime. The priming advantage in this condition has risen to 40 per cent. Why does the benefit of a related prime jump from 15 per cent to 40 per cent when the targets are more likely to be related to the primes? The answer is that, when there is a high chance that they will be related, participants spot the connection and try to guess what the target must have been: they often guess correctly. Notice that they can do this only because the prime word was clearly visible. Look now at the corresponding figures, for when the prime was displayed very briefly. Here the priming advantages (7 per cent and 9 per cent) are far more modest (but statistically significant). However, the important result is that the change from 10 per cent to 90 per cent relatedness does not produce the large increase in the priming effect observed in the 1 second condition. The small increase from 7 per cent to 9 per cent was not statistically significant. It can be concluded that participants were unable to

guess in the brief condition, so presumably had not been able to identify the prime words. Nevertheless, those words did produce a small priming effect, so they must have received sufficient analysis to activate their meaning.

## 2.2 Towards a theory of parallel processing

When people are asked to guess about masked material, they are commonly able to provide some information, but it often lacks detail. For example, if participants in a Sperling-type experiment have recalled three letters, but are pressed for more, then they can often provide one or two. However, they generally do not know information such as whereabouts in the display the letters occurred, or what colour they were. These, of course, are exactly the kinds of detail that can be used to select items for report, and were believed to be usable in that role because they were characteristics which could be processed quickly and in parallel. The guessing results seem to turn the logic on its head, because the presumed complex information, such as letter identities, is discovered, while the simple colour and position information is unavailable. Coltheart (1980) offered an elegant solution to this problem, built around the semantic/episodic distinction used when describing memory (see Chapter 8). In the context of letters, semantic information would be the basic knowledge of letter identity. Episodic detail links the general identity to a specific occurrence: detail such as the fact that 'N' is in large, upper-case type, and is printed in red and at the start of the sign 'NO SMOKING'. Coltheart proposed that items do not normally reach conscious awareness unless both the semantic and episodic detail are detected. So, for example, one would not expect to be having an 'N-feeling' (semantic) in the absence of a letter with some specific characteristics (size, colour, etc.) in the field of view!

It has become clear from electrophysiological studies that visual item identification occurs in a different region of the cortex from the areas which respond to colour or location. These different kinds of information have to be united, and this process, Coltheart (1980) suggests, takes time and attention. According to this account, Sperling's 12 letters, or even Evett and Humphrey's *lion*, are indeed processed in parallel to cause semantic activation, but the viewer will not become aware of this, unless able to assign the corresponding episodic details. Nevertheless, if pressed, the participant may sometimes admit to 'having a feeling' that an item might have been presented, although not know what it looked like (see also Chapter 8 for a discussion of the semantic–episodic distinction).

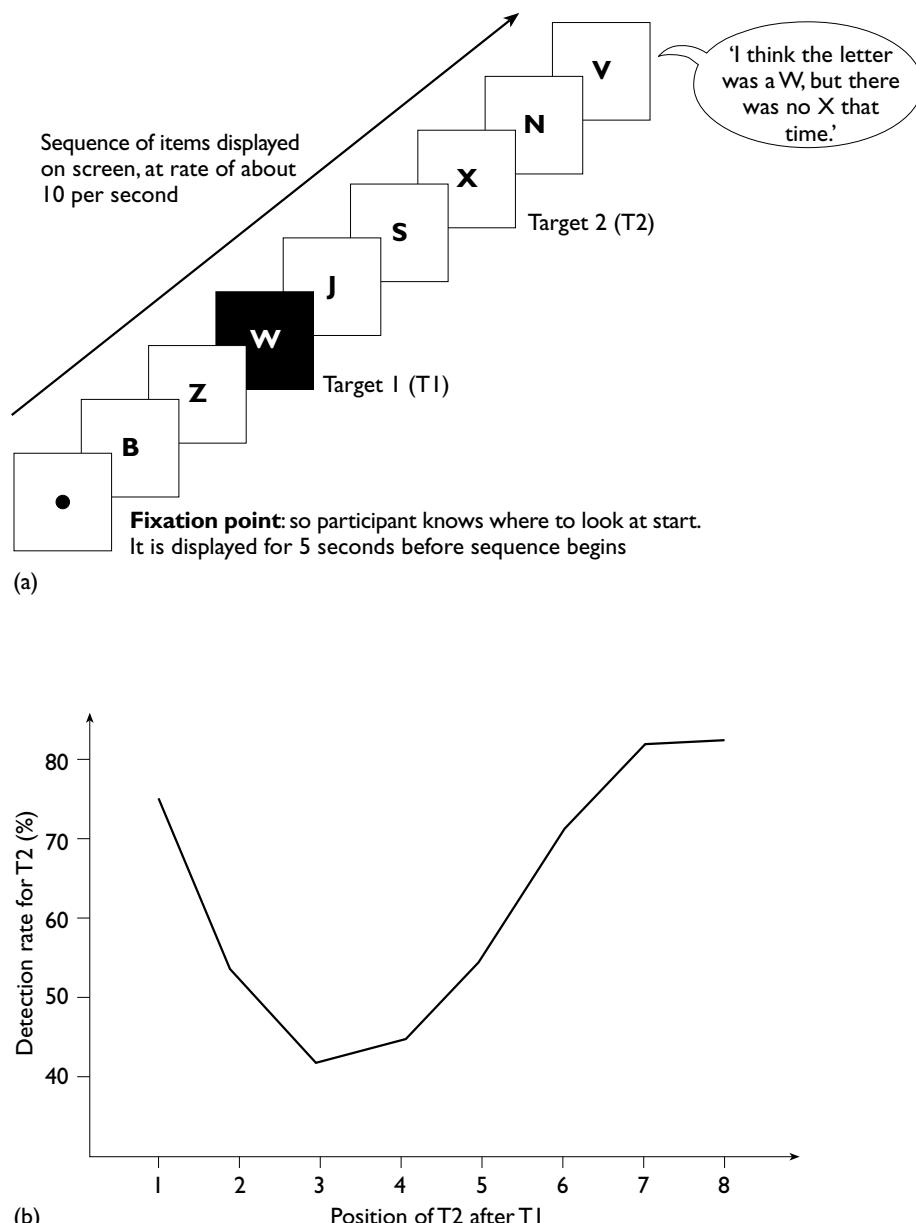
The important point to note in the above account is that attention is no longer being described as the process that selects material for complex serial processing (e.g. word identification). Instead, Coltheart suggests that attention is required to join the products of two parallel processes: the identification and the episodic characterization. This idea that attention is concerned with uniting the components of a stimulus is not unlike a theory which Treisman has been developing (after her early auditory attention work, she now researches visual attentive processes). We shall consider Treisman's work (which does not involve backward masking), but first we should look a little further at what masking actually does to the processing of a stimulus.

## 2.3 Rapid serial visual presentation

It has been known for a long time that backward masking can act in one of two ways: **integration** and **interruption** (Turvey, 1973). When the SOA between target and mask is very short, integration occurs; that is, the two items are perceived as one, with the result that the target is difficult to report, just as when one word is written over another. Of more interest is masking by interruption, which is the type we have been considering in the previous section. It occurs at longer SOAs, and interruption masking will be experienced even if the target is presented to one eye and the mask to the other. This dichoptic (two-eyed) interaction must take place after information from the two eyes has been combined in the brain; it could not occur at earlier stages. In contrast, integration masking does not occur dichoptically when target and mask are presented to separate eyes, so presumably occurs quite early in analysis, perhaps even on the retina. On this basis, Turvey (1973) described integration as peripheral masking, and interruption as central masking, meaning that it occurred at a level where more complex information extraction was taking place.

Another early researcher in the field (Kolers, 1968) described the effect of a central (interruption) mask by analogy with the ‘processing’ of a customer in a shop. If the customer (equivalent to the target) comes into the shop alone, then s/he can be fully processed, even to the extent of discussing the weather and asking about family and holidays. However, if a second customer (i.e. a mask) follows the first, then the shopkeeper has to cease the pleasantries, and never learns about the personal information. The analogy was never taken further, and of course it is unwise to push an analogy too far. Nevertheless, one is tempted to point out that the second customer is still kept waiting for a while. Where does that thought take us? It became possible to investigate the fate of following stimuli, in fact whole queues of stimuli, with the development of a procedure popularized by Broadbent (Broadbent and Broadbent, 1987), who, like Treisman, had moved on from auditory research. The procedure was termed Rapid Serial Visual Presentation, in part, one suspects, because that provided the familiar abbreviation RSVP; participants were indeed asked to *répondez s'il vous plaît* with reports of what they had seen.

Unlike the traditional two-stimulus, target/mask pairing, **Rapid Serial Visual Presentation (RSVP)** displayed a series of stimuli in rapid succession, so each served as a backward mask for the preceding item. SOAs were such that a few items could be reported, but with difficulty. Typical timings would display each item for 100 ms, with a 20 ms gap between them; the sequence might contain as many as 20 items. Under these conditions stimuli are difficult to identify, and participants are certainly unable to list all 20; they are usually asked to look out for just two. In one variation, every item except one is a single black letter. The odd item is a white letter, and this is the first target; the participant has to say at the end of the sequence what the white letter had been. One or more items later in the sequence (i.e. after the white target), one of the remaining black letters may be an ‘X’. As well as naming the white letter, the participant has to say whether or not X was present in the list. These two targets (white letter and black X) are commonly designated as T1 and T2. Notice that the participant has two slightly different tasks: for T1 (which will certainly be shown) an unknown letter has to be identified, whereas for T2 the task is simply to say whether a previously designated letter was presented. These details, together with a graph of typical results, are shown in Figure 2.3.



**Figure 2.3** The RSVP technique: (a) The sequence of stimuli, shown in the same location on a computer screen, in which the participant has to identify a white letter, then decide whether an X was also present; (b) Typical results, showing the likelihood of detecting the X, when presented in the first and subsequent positions following the white target

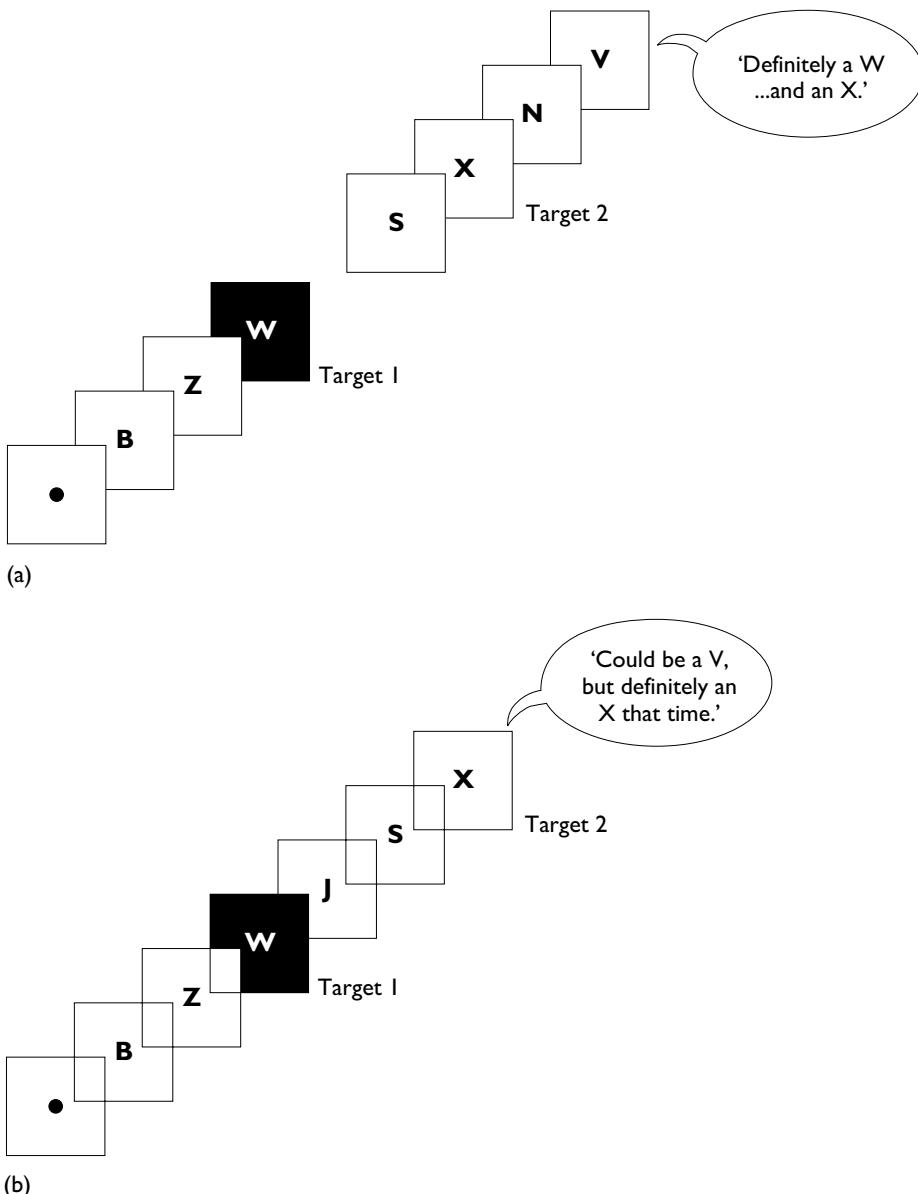
As can be seen from the graph in Figure 2.3(b), T2 (the X) might be spotted if it is the item immediately following T1, but thereafter it is less likely that it will be detected unless five or six items separate the two. What happens when it is not detected? As you may be coming to expect, the fact that participants do not report T2 does not mean that they have not carried out any semantic analysis upon it. Vogel *et al.* (1998) conducted an RSVP experiment that used words, rather than single letters. Additionally, before a sequence of stimuli was presented, a clear ‘context’

word was displayed, for a comfortable 1 second. For example, the context word might be *shoe*, then the item at T2 could be *foot*. However, on some presentations T2 was not in context; for example, *rope*. While participants were attempting to report these items, they were also being monitored using EEG (electro-encephalography). The pattern of electrical activity measured via scalp electrodes is known to produce a characteristic ‘signature’, when what might be called a mismatch is encountered. For example, if a participant reads the sentence *He went to the café and asked for a cup of tin*, the signature appears when *tin* is reached. The Vogel *et al.* (1998) participants produced just such an effect with sequences such as *shoe – rope*, even when they were unable to report seeing *rope*. This sounds rather like some of the material discussed earlier, where backward masking prevented conscious awareness of material that had clearly been detected. However, the target in the RSVP situation appears to be affected by something that happened *earlier* (i.e. T1), rather than by a following mask. The difference needs exploring and explaining.

Presumably something is happening as a result of processing the first target (T1), which temporarily makes awareness of the second (T2) very difficult. Measurements show that for about 500 to 700 ms following T1, detection of T2 is lower than usual. It is as if the system requires time to become prepared to process something fresh, a gap that is sometimes known as a **refractory period**, but that in this context is more often called the **attentional blink**, abbreviated to AB. While the system is ‘blinking’ it is unable to attend to new information.

Time turns out not to be the only factor in observing an AB effect ('AB effect' will be used as a shorthand way of referring to the difficulty of reporting T2). Raymond *et al.* (1992) used a typical sequence of RSVP stimuli, but omitted the item immediately following the *first* target. In other words, there was a 100 ms gap, rather than another item following. Effectively, this meant that the degree of backward masking was reduced, and not surprisingly resulted in some improvement in the report rate for T1. Very surprisingly, it produced a considerable improvement in the reporting of T2; the AB effect had vanished (see Figure 2.4(a)). How did removing the mask for one target lead to an even larger improvement for another target that was yet to be presented? To return to our earlier analogy, if the shopkeeper is having some trouble in dealing with the first customer, then the second is kept waiting and suffers. That doesn't explain *how* the waiting queue suffers (if it were me I should probably chat to the person behind, and forget what I had come for), but that question was also addressed by removing items from the sequence.

Giesbrecht and Di Lollo (1998) removed the items following T2, so that it was the last in the list; again, the AB effect disappeared (see Figure 2.4(b)). So, no matter what was going on with T1, T2 could be seen, if it was not itself masked. To explain this result, together with the fact that making T1 easier to see also helps T2, Giesbrecht and Di Lollo developed a two-stage model of visual processing. At Stage 1, a range of information about target characteristics is captured in parallel: identity, size, colour, position and so on. In the second stage, they proposed, serial processes act upon the information, preparing it for awareness and report. While Stage 2 is engaged, later information cannot be processed, so has to remain at Stage 1. Any kind of disruption to T1, such as masking, makes it harder to process, so information from T2 is kept waiting longer. This has little detrimental impact upon T2 unless it



**Figure 2.4** (a) Target 2 is seen more easily when Target 1 is made easier to see by removing the following item; (b) Target 2 is also seen easily when items following it are omitted

too is masked by a following stimulus (I don't forget what I came to buy, if there is no-one else in the queue to chat with). When T2 is kept waiting it can be overwritten by the following stimulus. The overwriting process will be damaging principally to the episodic information; an item cannot be both white and black, for example. However, semantic information may be better able to survive; there is no reason why *shoe* and *rope* should not both become activated. Consequently, even when there is insufficient information for Stage 2 to yield a fully processed target, it may

nevertheless reveal its presence through priming or EEG effects. There is an obvious similarity between this account and Coltheart's (1980) suggestion: both propose the need to join semantic and episodic detail.

## 2.4 Masking and attention

Before I summarize the material in this section, and we move on to consider attentional processes with clearly-seen displays, it would be appropriate to consider the relevance of the masking studies to the issue of attention. We began the whole subject by enquiring about the fate of material which was, in principle, available for processing, but happened not to be at the focus of attention. Somehow we have moved into a different enquiry, concerning the fate of material that a participant was trying to attend to, but did not have time to process. This seemed a natural progression as the chapter unfolded, but are the two issues really related? Merikle and Joordens (1997) addressed this very question; they characterized it as a distinction between perception without awareness (such as in masking studies) and perception without attention (as with dichotic listening). They carried out a number of studies, in which processing was rendered difficult either by masking, or by giving the participants two tasks, so that they could not focus on the target. They concluded that the results were entirely comparable, and that the same underlying processes are at work in both kinds of study.

## Summary of Section 2

---

The results of the visual attention experiments we have considered can be interpreted as follows.

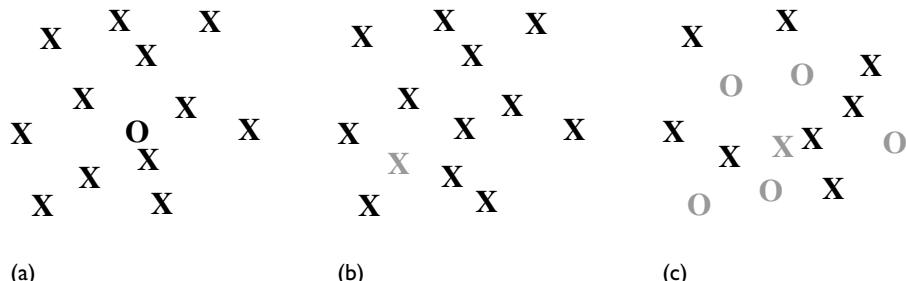
- Attention can be directed selectively towards different areas of the visual field, without the need to re-focus.
- The inability to report much detail from brief, masked visual displays appears to be linked to the need to assemble the various information components.
- The visual information is captured in parallel, but assembly is a serial process.
- Episodic detail (e.g. colour, position) is vulnerable to the passage of time, or to 'overwriting' by a mask.
- Semantic information (i.e. identity/meaning) is relatively enduring, but does not reach conscious awareness unless bound to the episodic information.
- Attention, in this context, is the process of binding the information about an item's identity to its particular episodic characteristics.
- 'Unbound' semantic activation can be detected by priming and electrophysiological techniques.

### 3 Integrating information in clearly-seen displays

The binding of features emerges as being a very significant process when displays are brief, because there is so little time in which to unite them. With normal viewing, such as when you examine the letters and words on this page, it is not obvious to introspection that binding is taking place. However, if, as explained above, it is a necessary precursor to conscious awareness, the process must also occur when we examine long-lived visual displays. Researchers have attempted to demonstrate that the binding process does indeed take place.

#### 3.1 Serial and parallel search

Examine the three sections of Figure 2.5 and in each case try to get a feel for how long it takes you to find the ‘odd one out’. The figure is a monochrome version of the usual form of these stimuli; you can see a coloured example in colour Plate 3.



**Figure 2.5** Find the odd item in each of the groups, (a), (b) and (c)

You probably felt that the odd items in Figures 2.5(a) and 2.5(b) simply ‘popped out’, and were immediately obvious, whereas the grey X in Figure 2.5(c) took you slightly longer to find. These kinds of effect have been explored formally by Treisman (e.g. Treisman and Gelade, 1980). The odd item is referred to as the target and the others as the distractors. Treisman showed her participants a series of displays of this nature, and measured how long it took them to decide whether or not a display contained a target. She was particularly interested in the effect of varying the number of distractors surrounding the targets. It was found that for displays similar to Figures 2.5(a) and 2.5(b) it made no difference to decision times whether there were few or many distractors. In contrast, with the 2.5(c) type of display, participants took longer to decide when there were more distractor items; each additional distractor added approximately 60 ms to the decision time.

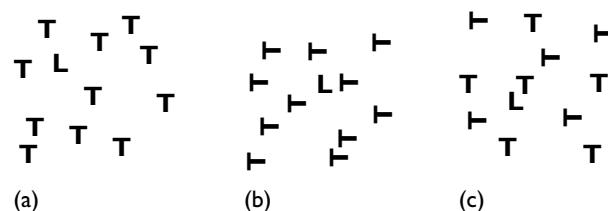
How is that pattern of results to be explained? Treisman pointed out that the first two displays have target items which differ from the rest on only one dimension; the target is either a round letter (O), among ‘crossed-line’ letters (X), or a grey letter among black letters. The 2.5(c) display type is different; to identify the target it is necessary to consider two dimensions. It has to be an X (but there are others, so on its own being an X does not define the target), and it has to be grey (but again, there are other grey letters). Only when X and grey are combined does it become clear that this is an ‘odd one out’. All these features (various colours and shapes) are quite simple and are derived in the early stages of visual processing, but importantly different

types of analysis (e.g. of shape or colour) take place in different parts of the brain. To see whether there is just ‘greyness’, or just ‘roundness’ in a display is easy, so easy in fact that the whole display seems to be taken in at a glance, no matter how many items there are. In other words, all the different items are processed at the same time, in parallel. The situation is very different when shape and colour have to be combined because they are determined in different brain areas; somehow the two types of information have to be brought together. You will recall from Section 2 that attention appears necessary to unite episodic and semantic information. Treisman proposed that it is also required to link simple features. Each item in the display has to receive attention just long enough for its two features (shape and colour) to be combined, and this has to be done one item at a time until the target is found. In other words, the processing is serial, so takes longer when there are more items to process.

It has been known for some time that the parietal region of the brain (part of the cortex that sits like a saddle across the top of the brain) is one of the areas involved in attention. A fuller account of the problems that result from damage to this area will be given in Section 5.1; at this point it is relevant to mention that Treisman (1998) reports investigations with a patient who had suffered strokes in that region. He was shown simple displays, containing just two letters from a set of three (T, X and O); they were printed in different colours, from a choice of three (red, blue or yellow). He was asked to describe the first letter he noticed in the display. On a particular occasion he might be shown a blue T and a red O. Although he often made mistakes, he would rarely respond ‘Yellow X’ to that display; that is, he did not claim to see features that were not there at all, so he was not simply guessing. What he did say quite often would be something like ‘Blue O’. He had correctly identified features that were present, but was unable to join them appropriately. The implication of this is that both the detection and the integration of features are necessary steps in normal perception, and that integration requires attention.

### 3.2 Non-target effects

Treisman’s **feature integration theory** has been very influential, but it does not appear to explain all experimental observations, and there have been alternative accounts of the feature-binding process. Duncan and Humphreys (1989) reported effects which do not fit too well within the basic Treisman account. They required participants to search for the letter ‘L’ (the target) within a number of ‘Ts’ (the non-targets). You may get a feel for the relative difficulty of different versions of their task by examining Figure 2.6.



**Figure 2.6** Examples of the kinds of stimuli used by Duncan and Humphreys (1989). Find the letter L in each of the groups, (a), (b) and (c)

The task can be conceptualized as looking for two lines that meet at a corner (the L), rather than forming a T-junction. It should not make much difference whether the T-junctions are vertical or horizontal (as in Figure 2.6(a) and 2.6(b)), and, indeed, the search times for these two sorts of display are similar. However, when the Ts are mixed, as in Figure 2.6(c), it takes longer to find the target. This finding would not have been predicted by a simple feature integration theory. Duncan and Humphreys (1989) argued that part of finding the target actually involves rejecting the non-targets and that this is a harder task when they come in a greater variety.

This explanation does not rule out the idea that features need to be integrated to achieve recognition, but it does suggest that non-targets, as well as targets, need to be recognized. The following section also describes evidence that non-targets are recognized, but in this case the recognition appears to take place in parallel.

### 3.3 The ‘flanker’ effect

A potential problem for the feature integration theory is the fact that the time taken to understand the meaning of a printed word can be influenced by other, nearby words. Of itself, this is not surprising, because it is well known that one word can prime (i.e. speed decisions to) another related word; the example *nurse – doctor* was given in Section 1.3. However, Shaffer and LaBerge (1979) found priming effects, even when they presented words in a way which might have been expected to eliminate priming. For their experiment a word was presented on a screen, and as quickly as possible a participant had to decide to what category it belonged; for example an animal or a vegetable. The participant was required to press one button for animal names, and another for vegetables. This sounds straightforward, but the target word was not presented in isolation; above and below it another word was also printed, making a column of three words. The target, about which a decision was to be made, was always in the centre. The words repeated above and below the target were termed the ‘flankers’. Before the three words were displayed, markers in the field of view showed exactly where the target would appear. Figure 2.7 shows examples of possible displays.

cat	pea
dog	dog
cat	pea
(a)	(b)

**Figure 2.7** The flanker effect. It takes longer to decide ‘dog’ is an animal when surrounded by words of another category, as in (b)

You will probably not be surprised to learn that people make category judgements more quickly for examples such as that shown in Figure 2.7(a) than for the 2.7(b) type of stimulus. Presumably, while the target information is being processed, details about the flankers are also being analysed, in parallel. When they turn out to be from the category associated with pressing the other button they slow the response. This slowing is very much like the impact of the conflicting colour names in the Stroop effect (see Box 2.2). However, recall that Treisman’s theory

suggests that focused, *serial* attention is required to join features together. A printed word has many features, and it would be thought that they require joining before the word can be recognized; it should not be possible to process the three words simultaneously. A participant focusing on the target could not (according to the theory) also be processing the flankers.

## 2.2

### Research study

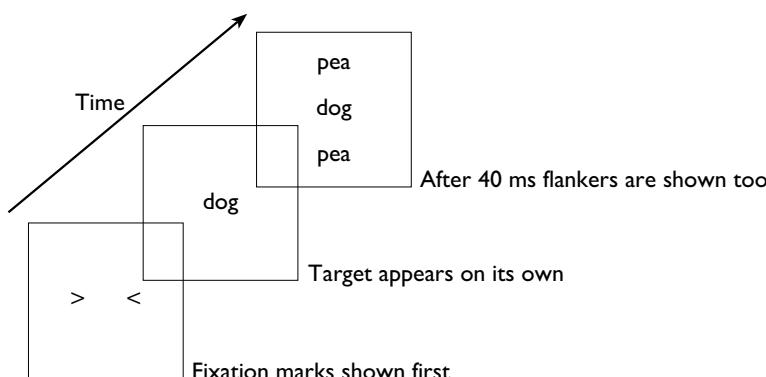
#### The Stroop effect

Stroop (1935) reported a number of situations in which the processing of one source of information was interfered with by the presence of another. The best known example uses a list of colour names printed in non-matching coloured inks (see Plate 4).

A variant is the ‘Emotional Stroop task’, which can be used in therapeutic diagnoses. For example, severe depression produces cognitive impairment and, in the elderly, it is difficult to distinguish this from the effects of the onset of dementia. Dudley *et al.* (2002) used colours to print a list of words, some of which were associated with negative emotions (e.g. the word *sadness*). Depressed people have an attentional bias towards such depression-related material. Patients were required to name the ink colours for each word, as quickly as possible. Both depressed patients and those in the early stages of Alzheimer’s disease were slower than a control group, but only the patients with depression were extra slow in responding to negative words. The technique permits an appropriate diagnosis.

Broadbent addressed this problem (Broadbent and Gathercole, 1990), and produced an explanation to ‘save’ the feature integration theory. He suggested that the central target word primed the flankers so effectively that they could be detected with the minimum of attention. Taking the items in Figure 2.7 as an example, if this explanation were true it would have to be argued that ‘dog’ primes ‘cat’, which, being another animal leads to faster decision times. ‘Dog’ cannot prime ‘pea’, as they are unrelated, so there is nothing to make the decision any quicker. In other words, it is not that ‘pea’ makes responses to ‘dog’ harder; rather, ‘cat’ makes them easier. Broadbent and Gathercole tested this explanation with an ingenious modification to the usual way of presenting targets and flankers. Instead of displaying all three words simultaneously, the target appeared first, to be joined by the flankers 40 ms later. The sequence is represented in Figure 2.8.

The reasoning behind this change was as follows. If Broadbent and Gathercole were correct that the flankers were analysed only because of priming from the target word, then giving the target a ‘head start’ should enable it to prime even more effectively; the flanker effect would be even stronger. On the other hand, if interference from the flankers were merely an example of processing not being as ‘serial’ as Treisman supposed, then making flankers arrive late, when target processing had already started, should *reduce* their impact. The results showed a strong flanker effect (i.e. faster responses with same-category flankers), suggesting that the priming idea was correct. However, there is another interpretation of the



**Figure 2.8** The Broadbent and Gathercole (1990) modification: the flankers are delayed for 40 ms

Broadbent and Gathercole results. It has been well established that an item suddenly appearing in the visual field will capture attention (e.g. Gellatly *et al.*, 1999). By making the flankers appear later, Broadbent and Gathercole may have ensured that they would attract attention away from the target. This could explain why the flankers showed a particularly strong effect with this style of presentation. Although the Broadbent and Gathercole idea of staggering the display times of the stimuli was ingenious, a convincing demonstration of parallel processing requires all the different stimuli to be presented at the same time.

## Summary of Section 3

---

When consciously perceiving complex material, such as when looking for a particular letter of a particular colour:

- Perception requires attention.
- The attention has to be focused upon one item at a time, thus ...
- processing is serial.
- Some parallel processing may take place, but ...
- it is detected indirectly, such as by the influence of one word upon another.

## 4 Attention and distraction

The above account of having attention taken away from the intended target reminds us that, while it may be advantageous from a survival point of view to have attention captured by novel events, these events are actually distractions from the current object of attention. Those who have to work in open-plan offices, or try to study while others watch TV, will know how distracting extraneous material can be. Some try to escape by wearing headphones, hoping that music will be less distracting, but does that work? Are some distractors worse than others? These kinds of question have been addressed by research and the answers throw further light upon the nature of attention.

## 4.1 The effects of irrelevant speech

Imagine watching a computer screen, on which a series of digits is flashed, at a nice easy rate of one per second. After six items you have to report what the digits had been, in the order presented (this is called serial recall – see also Chapters 9 and 16). Not a very difficult task, you might think, but what if someone were talking nearby? It turns out that, even when participants are instructed to ignore the speech completely, their recall performance drops by at least 30 per cent (Jones, 1999).

In the context of dichotic listening (Section 1.2), it was shown that ignored auditory material may nevertheless be processed, and hence its meaning influences perception of attended material. However, meaning appears to have no special impact, when speech interferes with memory for visually presented material. Thus, hearing numbers spoken, while trying to remember digits, is no more damaging than listening to other irrelevant speech items (Buchner *et al.*, 1996). In fact, even a foreign language, or English played backwards are no less disruptive than other irrelevant speech items (Jones *et al.*, 1990). On the other hand, simple white noise (a constant hissing like a mis-tuned radio) is almost as benign as silence. Interference presumably results from speech because, unlike white noise, it is not constant: it is broken into different sounds.

The importance of ‘difference’ in the speech can be shown by presenting lists of either rhyming or non-rhyming words. It turns out that a sequence such as ‘cat, hat, sat, bat ...’ is less disruptive than a sequence such as ‘cat, dog, hit, bus ...’ (Jones and Macken, 1995). Jones (1999) proposes that, whether listening to speech, music, or many other types of sound, the process requires the string of sounds to be organized into perceptual ‘objects’. To recognize an auditory object, such as a word or melody, requires that the segments of the stream of sounds are identified, and it is also necessary to keep track of the order of the segments. This ordering process, which occurs automatically, interferes with attempts to remember the order of visually presented items. When the sounds contain simple repetitions (as with the rhyming ‘at’ sound) the ordering becomes simpler, so the memory task is less disrupted. This was demonstrated in a surprising but convincing way by Jones *et al.* (1999). Their participants attempted to remember visually presented lists, while listening through headphones to a repeating sequence of three syllables, such as the letter names ‘k ... l ... m ... k ... l ... m’. These were disruptive, since the three letters have quite different sounds. The experimenters then changed the way in which the speech was delivered. The ‘l’ was played through both headphones, so sounded in the middle (see Section 1.2, Box 2.1), but the ‘k’ was played only to the left ear and the ‘m’ was heard in the right. This manipulation results in the perception of three ‘streams’ of speech, one on the left, saying ‘kay, kay, kay ...’, one in the middle, repeating ‘ell’, and the last on the right saying ‘em’. The significant point is that instead of hearing a continually changing sequence, the new way of playing *exactly the same sounds* results in them sounding like three separate sequences each of which never changes. Remarkably, the result is that they are no longer as disruptive to the visual recall task.

This section has taken the concept of attention into a new area. Previously we have seen it as a means of separating information, or of directing the assembly of different aspects of the attended item. In most of the earlier examples it has appeared that a great deal of processing can take place in parallel, although the results may not all reach conscious awareness. The impact of irrelevant speech shows that parallel

processing is not always possible. It seems to break down in this case because demands are made on the same process – the process that places items in a sequence. Here it would seem that we have a situation where there really is a ‘bottleneck’, of the sort envisaged in early theories of attention (see Sections 1.2 and 1.3).

What of trying to study with music? Undoubtedly, ‘Silence is Golden’, but if music is to be played, then my suggestion is that it should perhaps be something that changes very slowly, such as the pieces produced by some of the minimalist composers.

## 4.2 Attending across modalities

The preceding section raised the issue of attention operating (and to some extent failing) across two sensory modalities. By focusing on distraction we ignored the fact that sight and sound (and other senses) often convey mutually supporting information. A classic example is lip-reading. Although few of us would claim any lip-reading skills, it turns out that, particularly in noisy surroundings, we supplement our hearing considerably by watching lip movements. If attention is concerned with uniting elements of stimuli from within one sense, then we might expect it to be involved in cross-modal (i.e. across senses) feature binding too. In this section we will look briefly at one such process.

A striking example of the impact of visual lip movements upon auditory perception is found in the **ventriloquism effect**. This is most commonly encountered at the cinema, where the loudspeakers are situated to the side of the screen. Nevertheless, the actor’s voice appears to emanate from the face on the screen, rather than from off to the side. Driver (1996) demonstrated just how powerful this effect could be. He presented participants with an auditory task that was rather like shadowing in dichotic listening (Section 1.3) – only much harder! The two messages, one of which was to be shadowed, did not go one to each ear: they both came from the same loudspeaker, and were spoken in the same voice. To give a clue as to which was to be shadowed, a TV monitor was placed just above the loudspeaker, showing the face of the person reading the to-be-shadowed message. By lip-reading, participants could cope to some extent with this difficult task. Driver then moved the monitor to the side, away from the loudspeaker. This had the effect of making the appropriate message seem to be coming from the lips. Since the other message did not get ‘moved’ in this way, the two now *felt* spatially separate and, although in reality the sounds had not changed, the shadowing actually became easier!

These kinds of effects have further implications at a practical level. The use of mobile telephones while driving a car has been identified as dangerous, and the danger is not limited to the case where the driver tries to hold the phone in one hand and steer with the other. If a hands-free headset is used of the type which delivers sound via an earpiece to just one ear, the caller’s voice sounds as if it is coming from one side. Attending to this signal has the effect of pulling visual attention towards the lateral message, reducing the driver’s responsiveness to events ahead (Spence, 2002).

## Summary of Section 4

---

We have seen that attentive processes will ‘work hard’ to unite information into a coherent whole.

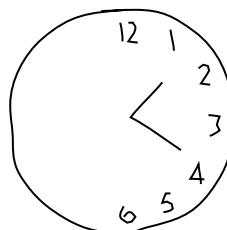
- Even spatially separate visual and auditory stimuli can be joined if they appear to be synchronous (the ventriloquism effect).
- When stimuli are not synchronous the system attempts to order the segments of the stimuli independently, resulting in distraction and lost information.
- It is a ‘bottleneck’ in the ordering process that results in one stream of information interfering with the processing of another.

## 5 The neurology of attention

Modern techniques for revealing where and when different parts of the brain become active have recently provided a window on the processes of attention. For example, one of these brain-scanning techniques, functional magnetic resonance imaging (fMRI), has been used to show the behaviour of an area of the brain that responds to speech. It turns out also to become activated in a person viewing lips making speech movements *in the absence of sound*. For this to happen there must be connections between relevant parts of the visual and auditory areas.

### 5.1 The effects of brain damage

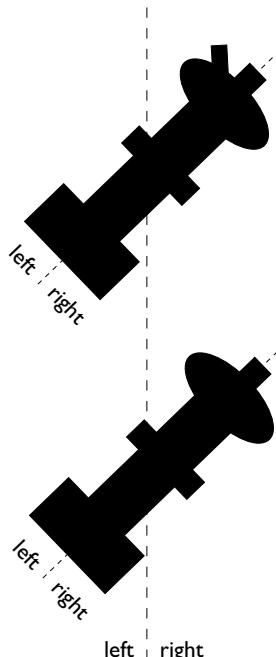
Before the advent of ‘brain mapping’, such as by fMRI, it was nevertheless possible to discover something of the part played by different regions of the brain, by observing the problems resulting from brain damage (such as following a stroke). One such area was mentioned in Section 3.1 – the parietal lobe. Damage to a single lobe (there is one on either side) leads to what is called **sensory neglect**, or sometimes simply neglect. A patient is likely completely to ignore the doctor if s/he stands on the neglected side (the side opposite to the site of the damage). When eating, the patient will probably leave any food that is on the ‘wrong’ side of the plate, and if asked to draw a flower will put petals on only one side. The problem is not simply blindness to all that lies on the neglected side. A patient asked to draw a whole vase of flowers may draw only those hanging over the ‘preserved’ side, but with each individual flower itself only half complete. It appears sometimes to be half the *object* which is neglected, rather than half the field of view. Figure 2.9 shows a typical attempt, by a patient with visual neglect, to draw a clock face.



**Figure 2.9** The typical appearance of a clock face, as drawn by a patient with visual neglect

That neglect may be associated with the object rather than the scene was demonstrated formally by Driver and Halligan (1991). They showed patients pairs of pictures that looked rather like silhouettes of chess pieces. Patients had to say whether the two pictures were the same or different. Where there *were* differences, they comprised an addition to one side, near the top of the figure (as if the chess queen had something attached to one ear!). When the addition was on the neglected side patients were unable to detect the difference. Suppose the ‘problem’ side was the left. The question is whether the patient has difficulty with processing information to the left of the page, or to the left of the object. Driver and Halligan tested this by tilting the pictures to the right (see Figure 2.10), so that the one-sided feature, although still on the left of the figure, was now in the right half of the page. Still the patients experienced difficulty: neglect was object-related.

We have been describing attention as a mechanism for assembling the sub-components of items in a scene, so it is not difficult to conceptualize a fault leading to some components being omitted. This account sees attention as an essential element of the perceptual process, helping to organize incoming information. However, neglect is not limited to objects that are physically present. Bisiach and Luzzatti (1978) asked their patient to imagine standing in the cathedral square of the Italian city where he grew up. He was to imagine looking towards the cathedral and to describe all that was in the square. He did this very well, except that he failed to mention any of the buildings down the left-hand side of the square (his brain injury was on the right). He was then asked to imagine standing on the cathedral steps, looking back towards his previous viewpoint. Again, he only reported details from the right. However, with the change of view, this meant that he was now describing



**Figure 2.10** Same or different? The feature that distinguishes the two figures is to the left of the object, but on the right of the page

previously ignored buildings! Clearly his memory was intact, but in some way not entirely accessible. Equally clearly, attentive processes are involved in the assembly of remembered material as well as of physically present stimuli.

An even more extreme form of neglect is encountered in a condition known as Balint's syndrome. It occurs when a patient is unfortunate enough to suffer damage to both parietal lobes, which results in it being extremely difficult to shift attention from one object to another. Thus, when trying to light a cigarette, the patient may find that his attention has been 'captured' by the flame, to the extent that he can no longer see the cigarette. One patient complained, 'When I see your spectacles I cannot see your face.' This is reminiscent of the experience of pilots using a head-up display (HUD) (see Section 2), where focusing on flight information displayed in the HUD makes the outside scene feel less 'visible'. Surprising as it may sound, it seems necessary to deduce from these effects that we *all* experience the world as a series of objects. However, unless our attentive process has been damaged, we can shift the attention so rapidly from one object to another that we perceive them all as being present simultaneously. Exactly what constitutes an object depends upon the situation; Balint patients are revealing here, because they see only one object at a time. Baylis *et al.* (1994) described a patient who could not report the letters making up an isolated word. Viewed in this way, each letter was a small object and it was not possible to switch attention from one to the next. However, the patient could read the whole word, since for this purpose it was a single object.

Early visual processing takes place in two major pathways in the brain, known as the ventral and dorsal streams (these are described in Section 6 of Chapter 3); the parietal region is part of the dorsal pathway. Damage to the ventral stream results in different kinds of integration problems; patients are aware of all aspects of a scene, but to the patient they remain segmented into small elements. For example, an individual shown a photograph of a paint-brush described seeing a wooden stick and a black object (the bristles) which he could not recognize. Humphreys (2001) suggests that the varieties of different problems are evidence that the binding together of different features takes place in several different stages and brain locations.

## 5.2 Event-related potentials

When a sense organ (eye, ear, etc.) receives a stimulus, the event eventually causes neurons to 'fire' (i.e. produce electrical discharges) in the receiving area of the brain. The information is sent on from these first sites to other brain areas. With appropriate apparatus and techniques it is possible to record the electrical signals, using electrodes attached to the scalp. The electrical potentials recorded are called **event-related potentials (ERPs)**, since they dependably follow the triggering sensory event. In fact a whole series of electrical changes are detected, first from the receiving brain areas, then later from subsequent sites. The timing of the ERPs gives a clue as to where in this sequence they are being generated.

Woldorff *et al.* (1993) examined ERPs evoked by sounds. These included signals occurring as soon as 10 ms after the auditory event. To generate a response so quickly, these ERPs must have originated in the brain stem, in the first 'relay' between ear and auditory cortex. The earliest stages of registration at the auditory cortex were detected after about 20–50 ms. It was of particular interest that, whereas the 10 ms signal was not affected by attention, the magnitude of the electrical activity

in the cortex was smaller when the sounds were played to an unattended ear. This shows that, at a very early stage of cortical analysis, attending away from a stimulus actually reduces the intensity of the signal in the brain. The result lends a good deal of support to the theory that attention is exercised by controlling a filter early in the processing sequence (see Section 1.3). Note, however, that the unattended signal is only attenuated, not eliminated.

## Summary of Section 5

---

Many familiar themes have re-emerged in this section, together with the recognition that attention is involved in the assembly of remembered material as well as of current perceptions.

- Attention is associated with the generation of perceptual objects.
- In addition to being an essential part of external stimulus processing, attention influences remembered experiences.
- ERP data show that cortical signals derived from unattended external stimuli are attenuated.

## 6 Concluding thoughts

We seem to have come a long way and covered a great deal of ground since I approached this subject by explaining that a mechanism must exist to help us focus on one sound out of many. That clearly is one function of attention, but attention seems to have other functions too. The results of visual search experiments show that attention is a vital factor in joining together the features that make up an object, and the experiences of brain-damaged patients suggest that this feature-assembly role ensures that our conscious perceptions are generally of objects, rather than of their constituent parts. Cross-modal research has demonstrated that the gathering together of related information from different senses is also controlled by attention.

Attention has a role to play in dealing with competition. The early researchers believed that attention was vital, because the brain would be able to deal with only one signal at a time; a ‘winning’ signal had to be picked from among the competitors. Although we have shown that a good deal of analysis can actually take place in parallel, there are also results which suggest that more complex analysis is largely serial, thus requiring a mechanism to select from the competing stimuli. Often, the parallel processes have to be demonstrated rather obliquely, since their results do not become consciously available. Thus attention has to do with what reaches conscious awareness. Why should this be so? Why should we not be equally aware of several items simultaneously?

Allport (1987) offered an answer that suggests yet another role for attention: it is to direct actions. Although we might, in principle, be able to perceive many things at once, there are situations where it would be counterproductive to attempt to *do* more than one thing. Allport gave fruit-gathering as an example. When we look at a bush of berries we need to focus attention upon one at a time, since that is how they have

to be picked. If animals had not evolved this ability to select, if all the food items remained equally salient, they would starve as they hovered over them all, unable to move toward any one! From this perspective, attention is the process that saves us from trying to carry out incompatible actions simultaneously. However, everyday experience reminds us that the issue of consciousness remains relevant. For example, novice drivers experience considerable difficulty in trying simultaneously to perform all the actions needed to control a vehicle; in Allport's view they are trying to 'attend-for-action' to more than one thing at a time. However, this could be restated as an attempt to be *conscious* of more than one thing at a time. Once the driver has become more skilful, the difficulty of combining actions disappears, but so too does the driver's conscious awareness of performing them: they have become automatic.

## 2.3

## Research study

### Hypnosis, time and attention

Brain scanning has revealed that regions of the brain known to be involved in attention show unusual activity when hypnotized participants become tolerant of pain (Crawford *et al.*, 1998), or experience hallucinations (Szechtman *et al.*, 1998).

Many people are unable to achieve such extreme effects in hypnosis, but there is one phenomenon that almost everyone experiences: hypnosis sessions usually feel to have lasted for far less time than the actual duration. I have explained this observation (Naish 2001, 2002) by linking it to Gray's (1995) theory of consciousness, which involves some of the same brain regions. He proposed that we maintain the content of our conscious awareness by registering repeated 'snapshots' of our environment. Our sense of time may be linked to the rate at which the environment is sampled.

To become hypnotized usually involves an induction in which one is asked to relax and focus attention on internal feelings, such as the heaviness of limbs or the rate of one's breathing. Subsequently, one is invited to imagine and attend to a pleasant, relaxing scene. Neither of these activities produces fast-changing streams of stimuli; the bodily feelings change only slowly and the relaxing scene is self-generated, so changes only when one wants it to change. I propose that in these circumstances there is no need to take such frequent snapshots, since little will change from one to the next. Consequently, we are less aware of the passage of time. In support of this claim, it turns out that participants who rate themselves as more successful at attending to their self-generated experiences and ignoring the real world are those who make larger underestimates of the session duration (Naish, 2003).

One might well ask how the term 'attention' has come to be applied to so many roles and processes; it might have been better to use different labels to distinguish between them. To use one word with so many aspects certainly makes a unitary definition very difficult to formulate. I suspect that the single term has stuck because ultimately all these facets of attention do lead to one result: conscious awareness. Even in so-called altered states of consciousness, such as hypnosis, attention appears

to be a vital component (see Box 2.3). To conclude with a personal view, I will offer the following definition:

*Attention is the process which gives rise to conscious awareness.*

I promised at the start of this chapter that attention was a broad and intriguing topic. I am sure you will agree that it was broad – and we haven’t covered half of it – but I hope you are now intrigued too. It is generally accepted that readers cannot continue to devote attention to text that goes on too long, so I trust that I have stimulated, rather than sated, your attention!

## Further reading

- Styles, E.A. (1997) *The Psychology of Attention*, Hove, Psychology Press. A very readable textbook, which covers and extends the topics introduced in this chapter.
- Pashler, H. (ed.) (1998) *Attention*, Hove, Psychology Press. An edited book, with contributors from North America and the UK. Topics are dealt with in rather more depth than in the Styles book.

## References

- Allport, D.A. (1987) ‘Selection for action: some behavioural and neurophysiological considerations of attention and action’, in Heuer, H. and Sanders, A.F. (eds) *Perspectives on Perception and Action*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Baylis, G.C., Driver, J., Baylis, L. and Rafal, R.D. (1994) ‘Reading of letters and words in a patient with Balint’s syndrome’, *Neuropsychological*, vol.32, pp.1273–86.
- Bisiach, E. and Luzzatti, C. (1978) ‘Unilateral neglect of representational space’, *Cortex*, vol.14, pp.129–33.
- Broadbent, D.E. (1952) ‘Listening to one of two synchronous messages’, *Journal of Experimental Psychology*, vol.44, pp.51–5.
- Broadbent, D.E. (1954) ‘The role of auditory localization in attention and memory span’, *Journal of Experimental Psychology*, vol.47, pp.191–6.
- Broadbent, D.E. and Broadbent, M.H.P. (1987) ‘From detection to identification: response to multiple targets in rapid serial visual presentation’, *Perception and Psychophysics*, vol.42, pp.105–13.
- Broadbent, D.E. and Gathercole, S.E. (1990) ‘The processing of non-target words: semantic or not?’, *Quarterly Journal of Experimental Psychology*, vol.42A, pp.3–37.
- Buchner, A., Irmens, L. and Erdfelder, E. (1996) ‘On the irrelevance of semantic information for the “Irrelevant Speech” effect’, *Quarterly Journal of Experimental Psychology*, vol.49A, pp.765–79.
- Cheesman, J. and Merikle, P.M. (1984) ‘Priming with and without awareness’, *Perception and Psychophysics*, vol.36, pp.387–95.
- Coltheart, M. (1980) ‘Iconic memory and visible persistence’, *Perception and Psychophysics*, vol.27, pp.183–228.

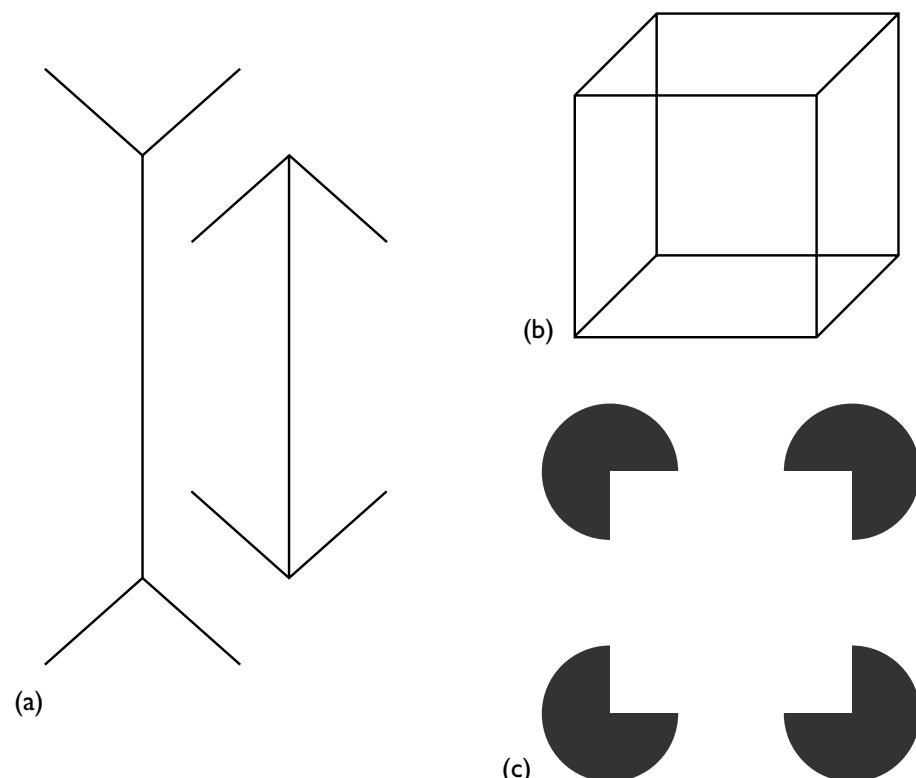
- Corteen, R.S. and Wood, B. (1972) 'Autonomous responses to shock associated words in an unattended channel', *Journal of Experimental Psychology*, vol.94, pp.308–313.
- Crawford, H.J., Horton, J.E., Hirsch, T.B., Harrington, G.S., Plantec, M.B., Vendemia, J.M.C., Shamro, C., McClain-Furmanski, D. and Downs, J.H. (1998) 'Attention and disattention (hypnotic analgesia) to painful somatosensory TENS stimuli differentially affects brain dynamics: a functional magnetic resonance imaging study', *International Journal of Psychophysiology*, vol.30, p.77.
- Deutsch, J.A. and Deutsch, D. (1963) 'Attention: some theoretical considerations', *Psychological Review*, vol.70, pp.80–90.
- Driver, J. (1996) 'Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading', *Nature*, vol.381, pp.66–8.
- Driver, J. and Halligan, P.W. (1991) 'Can visual neglect operate in object centred co-ordinates? An affirmative case study', *Cognitive Neuropsychology*, vol.8, pp.475–96.
- Dudley, R., O'Brien, J., Barnett, N., McGuckin, L. and Britton, P. (2002) 'Distinguishing depression from dementia in later life: a pilot study employing the emotional Stroop task', *International Journal of Geriatric Psychiatry*, vol.17, pp.48–53.
- Duncan, J. and Humphreys, G.W. (1989) 'Visual search and visual similarity', *Psychological Review*, vol.96, pp.433–58.
- Evett, L.J. and Humphreys, G.W. (1981) 'The use of abstract graphemic information in lexical access', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.33A, pp.325–50.
- Gellatly, A., Cole, G. and Blurton, A. (1999) 'Do equiluminant object onsets capture visual attention?', *Journal of Experimental Psychology: Human Perception and Performance*, vol.25, pp.1609–24.
- Giesbrecht, B. and Di Lollo, V. (1998) 'Beyond the attentional blink: visual masking by object substitution', *Journal of Experimental Psychology: Human Perception and Performance*, vol.24, pp.1454–66.
- Gray J.A. (1995) 'The contents of consciousness – a neuropsychological conjecture', *Behavioural and Brain Sciences*, vol.18, pp.659–76.
- Humphreys, G.W. (2001) 'A multi-stage account of binding in vision: neuropsychological evidence', *Visual Cognition*, vol.8, pp.381–410.
- Jones, D.M. (1999) 'The cognitive psychology of auditory distraction', *British Journal of Psychology*, vol.90, pp.167–87.
- Jones, D.M. and Macken, W.J. (1995) 'Phonological similarity in the irrelevant speech effect: within- or between-stream similarity?', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.103–15.
- Jones, D.M., Miles, C. and Page, J. (1990) 'Disruption of reading by irrelevant speech: effects of attention, arousal or memory?', *Applied Cognitive Psychology*, vol.4, pp.89–108.
- Jones, D.M., Saint-Aubin, J. and Tremblay, S. (1999) 'Modulation of the irrelevant sound effect by organizational factors: further evidence from streaming by location', *Quarterly Journal of Psychology*, vol.52A, pp.545–54.

- Kolers, P.A. (1968) ‘Some psychological aspects of pattern recognition’, in Kolers, P.A. and Eden. M. (eds) *Recognizing Patterns*, Cambridge, MA, MIT Press.
- Merikle, P.M. and Joordens, S. (1997) ‘Parallels between perception without attention and perception without awareness’, *Consciousness and Cognition*, vol.6, pp.219–36.
- Naish, P.L.N. (1990) ‘Simulating directionality in airborne auditory warnings and messages’, in Life, M.A., Narborough-Hall, C.S. and Hamilton, W.I. (eds) *Simulation and the User Interface*, London and New York, Taylor and Francis.
- Naish, P.L.N. (2001) ‘Hypnotic time distortion: busy beaver or tardy time-keeper’, *Contemporary Hypnosis*, vol.18, pp.118–30.
- Naish, P.L.N. (2002) ‘Perceiving, misperceiving, and hypnotic hallucinations’, in Roberts, D. (ed.) (2002).
- Naish, P.L.N. (2003) ‘The production of hypnotic time-distortion: determining the necessary conditions’, *Contemporary Hypnosis*, in press.
- Norman, D.A. (1968) ‘Towards a theory of memory and attention’, *Psychological Review*, vol.75, pp.522–36.
- Pecher, D., Zeelenberg, R. and Raaijmakers, G.W. (2002) ‘Associative priming in a masked perceptual identification task: evidence for automatic processes’, *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.55A, pp.1157–73.
- Raymond, J.E., Shapiro, K.L. and Arnell, K.A. (1992) ‘Temporary suppression of visual processing in an RSVP task: an attentional blink?’, *Journal of Experimental Psychology: Human Perception and Performance*, vol.18, pp.849–60.
- Roberts, D. (ed.) (2002) *Signals and Perception: The Fundamentals of Human Sensation*, Basingstoke, Palgrave/The Open University.
- Shaffer, W.O. and LaBerge, D. (1979) ‘Automatic semantic processing of unattended words’, *Journal of Verbal Learning and Verbal Behaviour*, vol.18, pp.413–26.
- Spence, C. (2002) ‘Multisensory integration, attention and perception’, in Roberts, D. (ed.) (2002).
- Sperling, G. (1960) ‘The information available in brief visual presentations’, *Psychological Monographs*, 74 (Whole Number 498), 29.
- Stroop, J.R. (1935) ‘Studies of interference in serial verbal reactions’, *Journal of Experimental Psychology*, vol.18, pp.643–62.
- Szechtman, H., Woody, E., Bowers, K.S. and Nahmias, C. (1998) ‘Where the imaginal appears real: a positron emission tomography study of auditory hallucinations’, *Proceedings of the National Academy of Sciences*, vol.95, pp.1956–60.
- Treisman, A. (1960) ‘Contextual cues in selective listening’, *Quarterly Journal of Experimental Psychology*, vol.12, pp.242–8.
- Treisman, A. (1998) ‘Feature binding, attention and object perception’, *Philosophical Transactions of the Royal Society of London*, Series B, vol.353, pp.1295–1306.

- Treisman, A. and Gelade, G. (1980) 'A feature-integration theory of attention', *Cognitive Psychology*, vol.12, pp.97–136.
- Turvey, M.T. (1973) 'On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli', *Psychological Review*, vol.80, pp.1–52.
- Vogel, E.K., Luck, S.J. and Shapiro, K.L. (1998) 'Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink', *Journal of Experimental Psychology: Human Perception and Performance*, vol.24, pp.1656–74.
- Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D. and Bloom, F.E. (1993) 'Modulation of early sensory processing in human auditory cortex during auditory selective attention', *Proceedings of the National Academy of Sciences of the USA*, vol.90, pp.8722–6.

### 1 Introduction

If you have ever searched frantically for an object that turns out to have been right in front of you all along, then this chapter may make you feel better. For, as you will see, perception of even the simplest object is actually a very complex affair. So, next time you turn the house upside down looking for your keys and then find them in the first place you looked, remember that your brain is using extremely sophisticated processes, many of which are beyond even the most advanced computer programs available today (not that computer programs ever lose their keys!).



**Figure 3.1** Three visual phenomena: (a) Müller-Lyer illusion; (b) Necker cube; (c) Kanizsa's illusory square

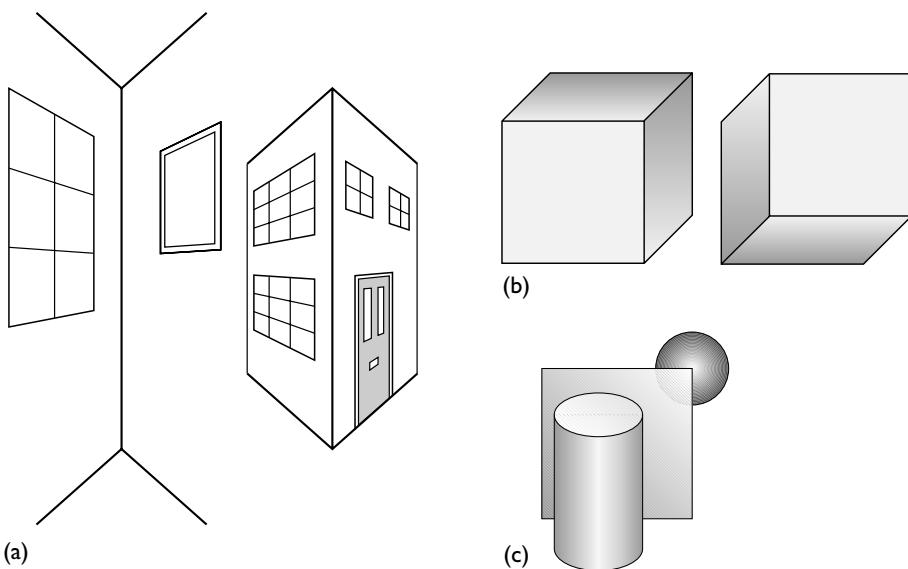
The sophistication of the cognitive processes that allow us to perceive visually is perhaps, if perversely, revealed best through the errors that our perceptual system can make. Figure 3.1 contains three very simple images that illustrate this. Image (a) is the Müller-Lyer illusion, in which the vertical line on the left is perceived as being longer even though both lines are of an identical length. Image (b) is a Necker cube, in which it is possible to perceive the cube in either of two perspectives (although you can never see both at the same time so please do not strain your eyes trying).

Image (c) is Kanizsa's (1976) illusory square, in which a square is perceived even though the image does not contain a square but only four three-quarter-complete circles.

### ACTIVITY 3.1

---

Look at each of the three visual illusions in Figure 3.1 and try to work out why it occurs. If you can't think of an answer, it may help to look at Figure 3.2.



**Figure 3.2** Some clues to why the illusions in Figure 3.1 may occur

#### COMMENT

One explanation for the Müller-Lyer illusion is that the arrowheads provide clues as to the distance of the upright line. For example, the inward-pointing arrowheads suggest that the vertical line might be the far corner of a room whilst the outward-pointing arrowheads suggest the vertical line could be the near corner of a building. We therefore see the first vertical line as longer because we assume it is further away from us than the second vertical line, though it makes the same size image on the retina.

The Necker cube can be seen in two different ways, as there are no clues as to which is the nearest face. Most cube-like objects that we encounter are solid and contain cues from lighting and texture about which is the nearest face. As the Necker cube does not contain these cues, we are unable to say for certain which face is closest.

Kanizsa's illusory square occurs due to a phenomenon known as perceptual completion. When we see an object partly hidden behind (occluded by) another object, we represent it to ourselves as a whole object rather than as missing its hidden parts. In the same way, we assume that four black circles are being occluded by a white square.

---

If the cognitive processes involved in perception were simple, then it would be hard to see how the effects in Figure 3.1 could occur. After all, they are all based on very straightforward geometric shapes that should be easy to perceive accurately. As we saw in Activity 3.1, there must be more sophisticated processes that have been developed to perceive the complex visual environment, which get confused or tricked by elements of these images. In fact the three effects above are likely to be caused because our visual system has evolved to perceive solid, three-dimensional (3D) objects and attempts to interpret the two-dimensional (2D) shapes as resulting from 3D scenes.

Perceptual errors arising from localized damage to the brain also demonstrate the complexities involved in visual perception. Some of the problems faced by people suffering from specific neuropsychological conditions include: being able to recognize objects but not faces (prosopagnosia); being able to perceive individual parts of the environment but not to integrate these parts into a whole; believing that one's family has been replaced by robots/aliens or impostors of the same appearance (Capgras syndrome); and only being able to perceive one side of an object (visual, or sensory, neglect – see Chapter 2, Section 5.1).

## 1.1 Perceiving and sensing

The term perception has different meanings, although a common element in most meanings is that perception involves the analysis of sensory information. When cognitive psychologists talk about perception, they are usually referring to the basic cognitive processes that analyse information from the senses. Throughout this chapter we shall be examining research and theories that have attempted to reveal and describe the cognitive processes responsible for analysing sensory information and providing a basic description of our environment; basically, how we make sense of our senses!

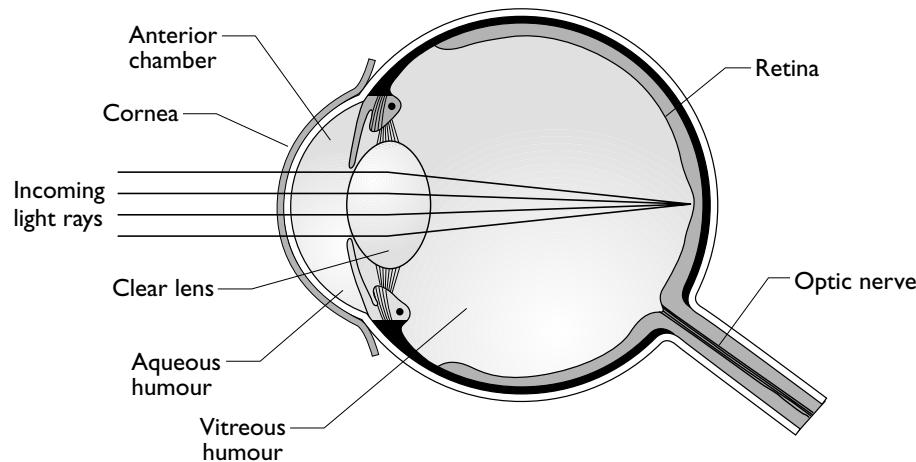
There has been considerable debate about the role played by sensory information in our perception of the world, with some philosophers rejecting the idea that it plays any part at all in the perception of objects. Atherton (2002) suggested that this may be because the notion of a sensation is rather problematic: 'Sensations seem to be annoying, extra little entities ... that somehow intervene between the round dish and our perception of it as round' (Atherton, 2002, p.4). We will not delve into this philosophical debate here, other than to note the distinction between sensation and perception. Throughout this chapter we will use the term 'sensation' to refer to the ability of our sense organs to detect various forms of energy (such as light or sound waves). However, to sense information does not entail making sense of it. There is a key difference between being able to detect the presence of a certain type of energy and being able to make use of that energy to provide information as to the nature of the environment surrounding us. Thus we use the term 'sensation' to refer to that initial detection and the term 'perception' to refer to the process of constructing a description of the surrounding world. For example, there is a difference between the cells in a person's eye reacting to light (sensation) and that person knowing that their course tutor is offering them a cup of tea (perception).

You may have noticed that we have begun to focus on visual perception rather than any of the other senses. Although the other senses, particularly hearing and touch, are undoubtedly important, there has been far more research on vision than on

the other modalities. This is because when we interact with the world we rely more on vision than on our other senses. Far more of the primate brain is engaged in processing visual information than in processing information from any of the other senses. We use vision both in quite basic ways, such as avoiding objects, and in more advanced ways, such as in reading or recognizing faces and objects. So, although the previous chapter examined auditory perception and Chapter 4 will explore haptic perception (touch) as well as visual perception, we will devote the present chapter to examining research into, and theories of, visual perception.

## 1.2 The eye

The logical place to start any consideration of visual perception is with the eye. A cross-section of the human eye is presented in Figure 3.3. Incoming light passes through the cornea into a small compartment called the anterior chamber (filled with fluid termed aqueous humour) and then through the lens into the major chamber of the eye that is filled with a viscous jelly called vitreous humour. The light is focused by the lens/cornea combination onto the retina on the back surface of the eye. It is the receptor cells in the retina that ‘sense’ the light.



**Figure 3.3** The human eye

The retina consists of two broad classes of receptor cell, rods and cones; so called for their shapes. Both rods and cones are sensitive to light, although the rods respond better than the cones at low light levels and are therefore the cells responsible for maintaining some vision in poor light. The cones are responsible for our ability to detect fine detail and different colours and are the basis of our vision at higher (daylight) light levels. Many animals, such as dogs and cats, have a higher ratio of rods to cones than humans do. This allows them to see better in poor light, but means that they are not so good at seeing either colour or fine detail.

One area of the retina that is of particular interest is the central portion known as the macula lutea (as it is yellow in colour and ‘lutea’ derives from a Latin word that means yellow), which contains almost all of the cones within the human retina. Within the macula, there is a small indentation called the fovea. The fovea is the area

of the retina that contains the highest density of cones and is responsible for the perception of fine detail.

### ACTIVITY 3.2

---

Place your thumbs together and hold them out at arm's length from your eyes. Now slowly move your left thumb to the left whilst keeping your eyes focused on your right thumb. You will find that after you have moved your left thumb more than about two thumb widths away from where your eyes are focused, that it appears to go out of focus. This is because the light being reflected into your eyes from the left thumb is no longer striking the fovea, meaning that you cannot perceive it in fine detail.

---

### 1.3 Approaches to perception

Psychologists have taken many different approaches to studying perception. One important distinction between approaches is whether the 'goal' of perception is assumed to be action or recognition. It is possible to conceive of recognition and action as being stages in the same perceptual process, so that action would only happen once recognition had taken place. However, our reaction to objects in the environment sometimes has to be very quick indeed, so that first having to work out what an object may be would be inconvenient to say the least. For example, if I see a moving object on a trajectory that means it will hit me in the head, the most important thing is to move my head out of the way. Working out that the object is the crystal tumbler containing vodka and tonic that was only moments ago in the hand of my somewhat angry looking partner is, for the moment at least, of secondary importance. I need to act to get out of the way of the object regardless of what the object actually is or who threw it.

As we shall see, there is evidence that perception for action and perception for recognition are quite different processes that may involve different neural mechanisms (Milner and Goodale, 1998). But, although it is important to make the distinction between perception for action and perception for recognition, we should not see them as being entirely independent. Sometimes the object that is about to hit your head could be the football that David Beckham has just crossed from the wing, requiring a very different response from that to the crystal tumbler.

Another way of differentiating approaches to perception is to consider the 'flow of information' through the perceptual system. To see what we mean by this phrase, try Activity 3.3.

### ACTIVITY 3.3

---

Consider these two scenarios:

- 1 A blindfolded student trying to work out what the unknown object they have been handed might be.
- 2 A blindfolded student searching for their textbook.

Imagine you are the blindfolded student. What strategies do you think you might employ to complete the above two tasks successfully? Can you identify any key differences in these strategies?

### COMMENT

A common strategy to employ for the first scenario is to try to build-up a 'picture' of the object by gradually feeling it. A common strategy to employ for the second scenario is to hold in your mind the likely shape and texture of the book and to search the environment for an object that shares these characteristics. The key difference between these scenarios is the direction in which information about the object is 'flowing', demonstrated by how the student's existing knowledge of what objects look like is being utilized. In the first scenario, information is flowing 'upward', starting with an analysis of the information derived from the senses (in this case via touch). In the second scenario, information is flowing 'downward', starting with the knowledge of what books tend to feel like.

---

So, in the case of touch, perception of the environment can involve information 'flowing' through the relevant perceptual system in two directions. But what about vision? If we were to remove the blindfold from our student in Activity 3.3, they would instantly be able to tell what the unknown object was or to spot the book in front of them. Does this mean that there is not a similar flow of information when the sense being used is vision?

To answer this question, let's try to formulate the stages involved in the student perceiving that there is a book in front of them. One approach might be:

- Light reflected from the book strikes the retina and is analysed by the brain.
- This analysis reveals four sudden changes in brightness (caused by the edges of the book against whatever is behind it).
- Two of these are vertical edges and two are horizontal edges (the left/right and top/bottom of the book).
- Each straight edge is joined (by a right angle at each end) to two others (to form the outline of the book).
- Within these edges is an area of gradually changing brightness containing many small, much darker areas (the white pages with a growing shadow toward the spine and the much darker words).
- A comparison of this image with representations of objects seen previously suggests that the object is an open book.

As this approach starts with the image formed on the retina by the light entering the eye and proceeds by analysing this pattern to gradually build up a representation of the object in view, we refer to it as involving **bottom-up processing**. This means that the flow of information through the perceptual system starts from the bottom – the sensory receptors – and works upward until an internal representation of the object is formed.

There is, however, another way of recognizing the book. It is very likely that the student has seen many books in the past and has a fair idea of what a book should

look like. This existing knowledge regarding book appearance could come in very useful in finding the textbook. Instead of building up a picture of the environment by analysing sensory information alone, it could be that the student uses existing knowledge of what books look like to find this particular book. For example, they might progress like this:

- I know that books are rectangular in shape and have light pages with dark words.
- I can see something in front of me that matches this description, so it must be a book.

The flow of information in this latter example has been reversed. The student started with existing knowledge regarding the environment and used this to guide their processing of sensory information. Thus the flow of information progressed from the top down as it started with existing knowledge stored in the brain, and we refer to it as involving **top-down processing**.

So both haptic and visual perceptual processes may operate both by building up a picture of the environment from sensory information and by using existing knowledge to make sense of new information. In other words, the flow of information through the perceptual system can be either bottom-up or top-down. These concepts will be explored throughout this chapter and we shall examine theories that concentrate on one or other of these processes and also look at how they might interact.

## Summary of Section 1

---

- Even the perception of simple images involves sophisticated cognitive processing, as demonstrated by visual illusions and neuropsychological disorders.
- We use the term **sensation** to refer to the detection of a particular form of energy by one of the senses and the term **perception** to refer to the process of making sense of the information sent by the senses.
- In the human eye the lens and cornea focus light onto the retina, which contains receptor cells that are sensitive to light.
- Perception can have different goals. The most common goals are perception for action and perception for recognition.
- The bottom-up approach to perception sees sensory information as the starting point, with perception occurring through the analysis of this information to generate an internal description of the environment.
- The top-down approach to perception involves making greater use of prior knowledge, with this guiding the perceptual process.

## 2 The Gestalt approach to perception

As with Chapter 2, we are going to examine the various approaches that have been taken to studying visual perception in a more or less historical order. One of the

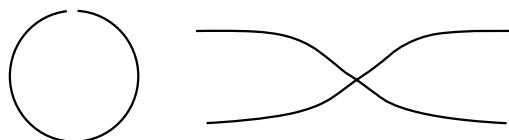
principal approaches to perception in the first half of the twentieth century was that of the **Gestalt** movement, which was guided by the premise ‘The whole is greater than the sum of its parts’. In perceptual terms, this meant that an image tended to be perceived according to the organization of the elements within it, rather than according to the nature of the individual elements themselves.

It is easy to see **perceptual organization** at work as it tends to be a very powerful phenomenon. In fact it appears as if both visual and auditory stimuli can be grouped according to similar organizing principles (Aksentijevic *et al.*, 2001).

### ACTIVITY 3.4

---

Look at Figure 3.4 and describe your first impression of what you see.



**Figure 3.4** Two examples of perceptual organization

---

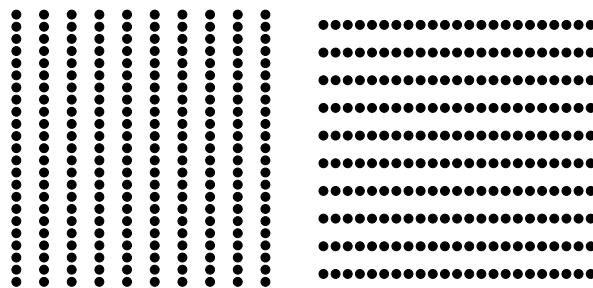
You probably described seeing a circle and two crossing lines. But, the image on the left is not a circle as it contains a gap at the top. This is the Gestalt perceptual organizational phenomenon of **closure** at work, in which a ‘closed’ figure tends to be perceived rather than an ‘open’ one. Likewise, the image on the right is not necessarily crossing lines, as it could be two pen-tips touching (in the middle of the image). The reason you see a cross is due to what the Gestalt researchers called **good continuation**, by which we tend to interpret (or organize) images to produce smooth continuities rather than abrupt changes.

Other Gestalt organizational laws, include **proximity** and **similarity**.

### ACTIVITY 3.5

---

As before, look at Figure 3.5 and describe your first impressions.



**Figure 3.5** The organizational law of proximity

---

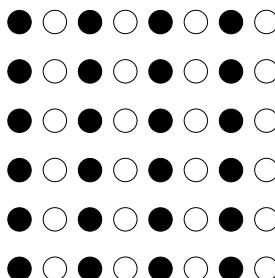
At one level you probably see two squares, due to the law of closure. However, you will also probably have seen the square on the left as consisting of columns of dots

and the one on the right as consisting of rows of dots. The reason for this is that, in the left-hand image, the horizontal spacing between the dots is greater than the vertical, and vice versa for the image on the right. Thus, the proximity of the individual elements is being used to group them into columns in the left-hand square and rows in the right-hand one.

### ACTIVITY 3.6

---

Now describe what you see in Figure 3.6.



**Figure 3.6** The organizational laws of similarity and proximity

---

As well as again seeing a square due to the law of closure, you perhaps saw the square as consisting of columns of circles. If so, this was an example of the organizational law of similarity (of colour). However, the spacing of the circles is such that the law of proximity encourages you to see rows not columns. For many people the law of similarity takes precedence and they see columns, while others may tend to see rows. Most people can readily switch between one organization (or interpretation) and the other because each conforms with a particular gestalt law.

The Gestalt researchers (including Koffka, 1935; Kohler, 1947 and Wertheimer, 1923) formulated other organizational laws, but most were deemed to be manifestations of the **Law of Pragnanz**, described by Koffka as: ‘Of several geometrically possible organizations that one will actually occur which possesses the best, simplest and most stable shape’ (Koffka, 1935, p.138).

So, you can see that a number of organizational laws can be used in order to work out which individual components of an image should be grouped together. Now look around the room in which you are sitting. How many squares composed of dots can you see? How many nearly complete circles and crossing lines are there? Your immediate response was probably to say ‘none’ or ‘only those in this book’. However, if you look carefully you will see that the stimuli used in the Gestalt demonstrations do have counterparts in the real world. For example, when I look out of my window I see a football that is partly hidden by a post and provides an example of closure, as I perceive a complete sphere rather than an incomplete circle. The figures that you have seen in this section can therefore be seen as simplified 2D versions of real-world objects and scenes. Because they are simplified, some information that would be present in real-world scenes is discarded. This lack of realism is a disadvantage. On the other hand, however, it is possible to control and

manipulate features of these figures, such as the proximity or similarity of elements, to see how they may contribute to perception.

As we shall see in the next section, there is considerable tension in the field of visual perception as to the usefulness of simplified stimuli such as those used by the Gestaltists. Some approaches are based on laboratory experimentation in which simplified scenes or objects are shown to participants, whilst proponents of other approaches claim that perception can only be studied in the real world, by examining how people perceive solid, 3D objects that are part of a complex 3D environment.

## Summary of Section 2

---

- The Gestalt approach to perception involved studying the principles by which individual elements tend to be organized together.
- Organizing principles include closure, good continuation, proximity and similarity.
- The stimuli used by Gestalt researchers tended to be quite simple, two-dimensional geometric patterns.

## 3 Gibson's theory of perception

In Section 1.3 we stated that one way of classifying different approaches to perception was according to whether they were primarily bottom-up or top-down. If visual perception is based primarily around bottom-up processing, we must be capable of taking the information from the light waves that reach our eyes and refining it into a description of the visual environment. Bottom-up perception requires that the light arriving at the retina is rich in information about the environment. One bottom-up approach to perception, that of J.J. Gibson (1950, 1966), is based on the premise that the information available from the visual environment is so rich that no cognitive processing is required at all. As Gibson himself said:

When the senses are considered as a perceptual system, all theories of perception become at one stroke unnecessary. It is no longer a question of how the mind operates on the deliverances of sense, or how past experience can organize the data, or even how the brain can process the inputs of the nerves, but simply how information is picked up.

*(Gibson, 1966, p.319)*

If you are thinking to yourself, ‘what does picked up mean?’ or ‘how is this information *picked up?*’, you are expressing a criticism that is often levelled at Gibson’s theory (e.g. Marr, 1982). The Gibsonian approach concentrates on the information present in the visual environment rather than on how it may be analysed. There is a strong link between perception and action in Gibson’s theory, and action

rather than the formation of an internal description of the environment can be seen as the ‘end point’ of perception.

Gibson conceptualized the link between perception and action by suggesting that perception is **direct**, in that the information present in light is sufficient to allow a person to move through and interact with the environment. One implication of this is that, whereas perception of a real environment is direct, perception of a 2D image in a laboratory experiment (or any 2D image come to that) would be *indirect*. When confronted with an image, our direct perception is that it is an image; that it is two-dimensional and printed on paper, for example. Our perception of that which it depicts is only indirect. For this reason, Gibson thought that perception could never be fully explored using laboratory experiments.



**Figure 3.7** *Ceci n'est pas une pipe*, 1928, by René Magritte

When you look at the Figure 3.7, what do you see? Your first reaction is probably to say ‘a pipe’. But, if what you are seeing is a pipe, then why can’t you pick it up and smoke it? As Magritte informs us, what you are seeing is not a pipe, but a picture of a pipe. Like Gibson, Magritte is drawing a distinction between direct perception (paint on canvas) and indirect perception (that the painting depicts a pipe).

### 3.1 An ecological approach

At the heart of Gibson’s approach to perception is the idea that the world around us structures the light that reaches the retina. Gibson believed perception should be studied by determining how the real environment structures the light that reaches our retina. From the importance placed on the ‘real world’ it is clear why Gibson’s is seen as an **ecological approach** to perception. Gibson referred to theories that were

based on experiments employing artificial, isolated, flat (or plane) shapes as ‘air’ theories, whilst he referred to his own as a ‘ground’ theory, as it emphasized the role played by the real, textured surface of the ground in providing information about distance. As Gibson stated: ‘A surface is substantial; a plane is not. A surface is textured; a plane is not. A surface is never perfectly transparent; a plane is. A surface can be seen; a plane can only be visualized’ (Gibson, 1979, p.35).

The impetus for Gibson’s theory came from his work training pilots to land and take-off during the Second World War. When approaching a runway, it is very important that a pilot is able to judge accurately the distance between plane and ground. The perceptual skill involved in this judgement is that of ‘depth perception’, this being the ability to judge how far you are from an object or surface. However, Gibson found that tests based on pictorial stimuli did not distinguish good from bad pilots and that training with pictorial stimuli had little impact on actual landing performance (Gibson, 1947). Extrapolating from this problem, Gibson suggested that psychological experimentation based on the use of pictorial stimuli is not an apt method for studying perception.

His point was that the experience of perception in the real world is very different from the experience of looking at 2D experimental stimuli in a laboratory. In the real world, objects are not set against a blank background, but against the ground, which consists of a very large number of surfaces that vary in their distance from and orientation to the observer. In their turn, these surfaces are not perfectly smooth planes, but consist of smaller elements, such as sand, earth and stone, which give them a textured appearance. In addition, the objects themselves will consist of real surfaces that also contain texture. To explain perception, we need to be able to explain how these surfaces and textures provide information about the world around us.

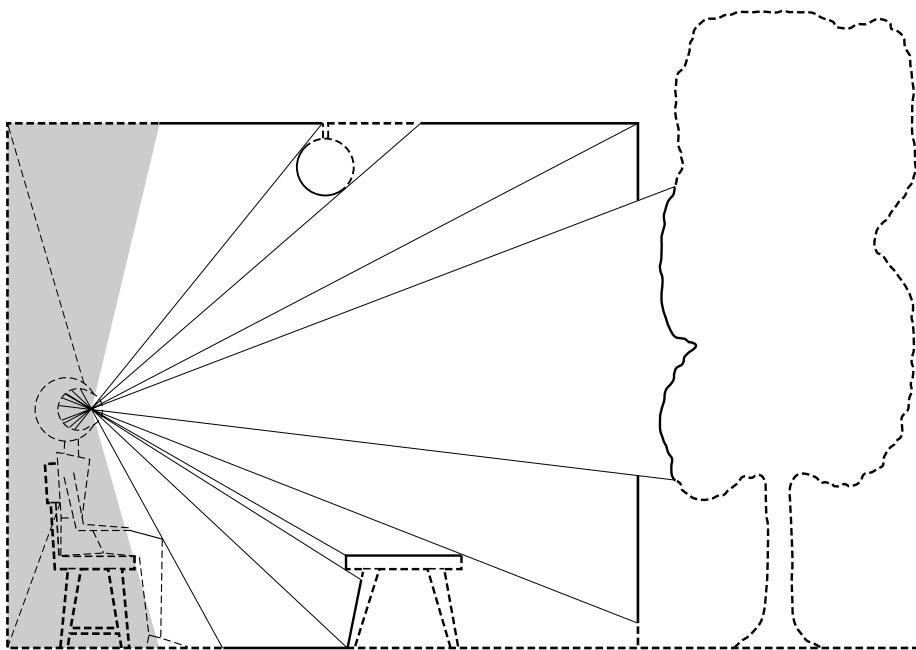
### 3.2 The optic array and invariant information

The structure that is imposed on light reflected by the textured surfaces in the world around us is what Gibson termed the **ambient optic array**. The basic structure of the optic array is that the light reflected from surfaces in the environment converges at the point in space occupied by the observer (see Figure 3.8). As you can see from Figure 3.9, as you stand up, the position of your head with respect to the environment is altered and the optic array changes accordingly.

You can see from Figures 3.8 and 3.9 that the primary structure of the optic array is a series of angles that are formed by light reflecting into the eyes from the surfaces within the environment. For example, an angle may be formed between the light that is reflected from the near edge of a table and that from the far edge.

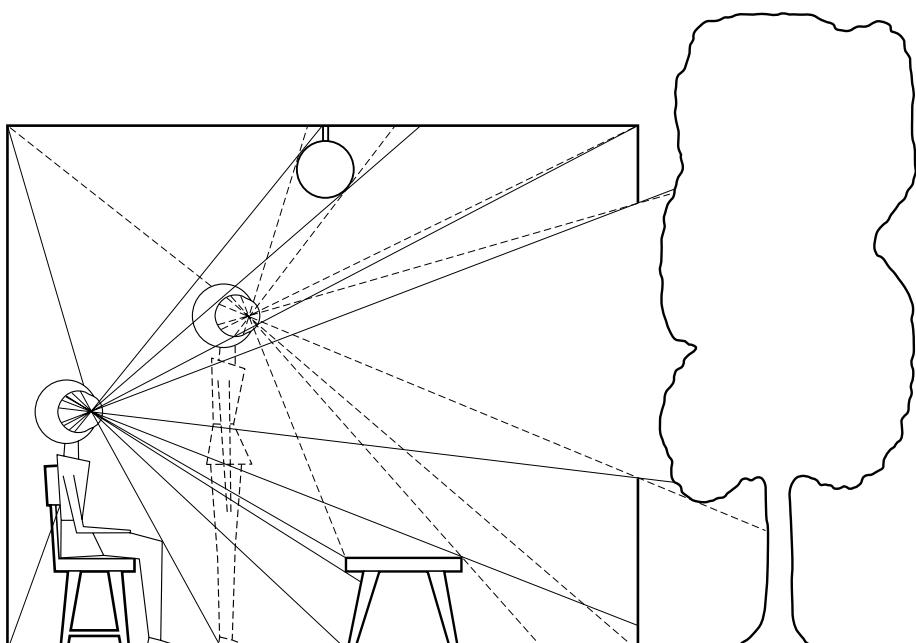
In addition to the primary structure of the optic array, Gibson maintained that there were additional, higher-order features that could provide unambiguous information as to the nature of the environment. He referred to these higher-order features as **invariants**, and believed that an observer could perceive the surrounding world by actively sampling the optic array in order to detect invariant information.

One of the most commonly cited forms of invariant information was explored by Sedgwick (1973). Sedgwick demonstrated the ‘horizon ratio relation’, which specifies that the ratio of how much of an object is above the horizon to how much is



**Figure 3.8** The ambient optic array

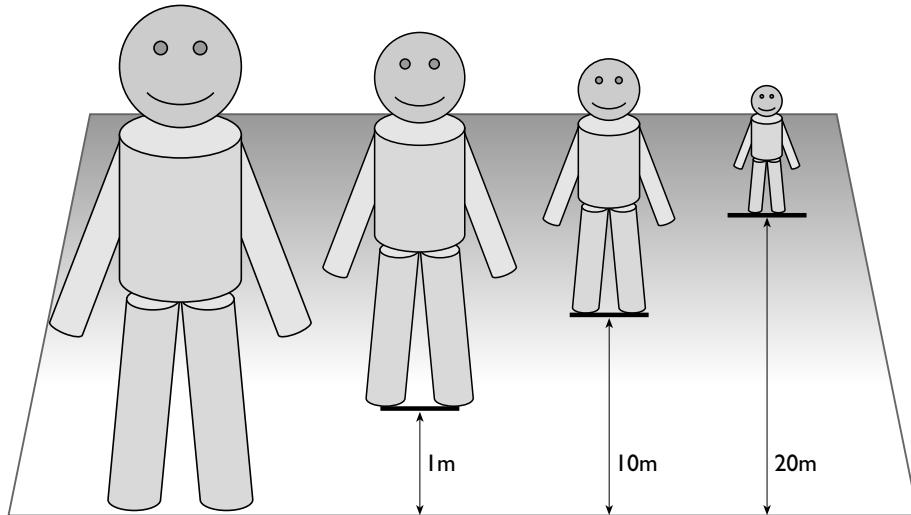
Source: Gibson, 1979, Figure 5.3



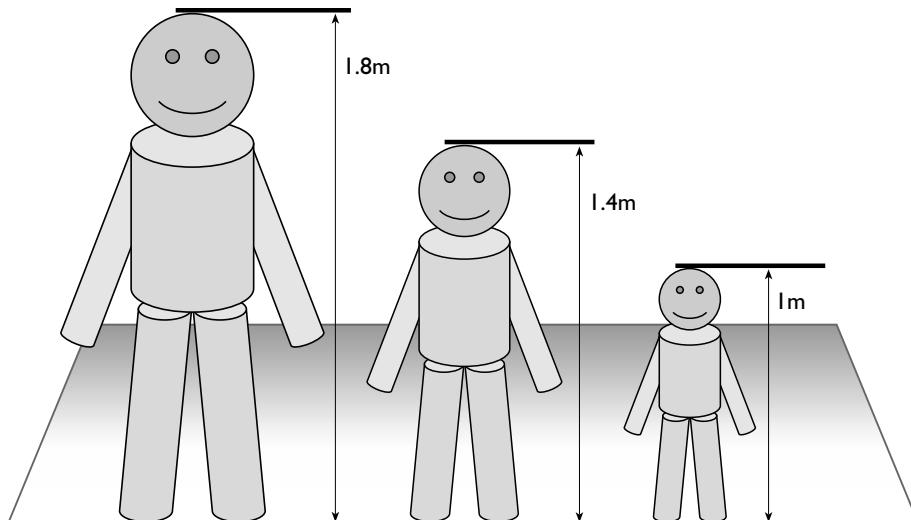
**Figure 3.9** Change in the optic array caused by movement of the observer

Source: Gibson, 1979, Figure 5.4

below remains constant (or invariant) as the object travels either toward or away from you (see Figure 3.10). This form of invariant information allows you to judge the relative heights of different objects regardless of how far away they are. The proportion of the object that is ‘above’ the horizon increases with the overall height of the object (see Figure 3.11).



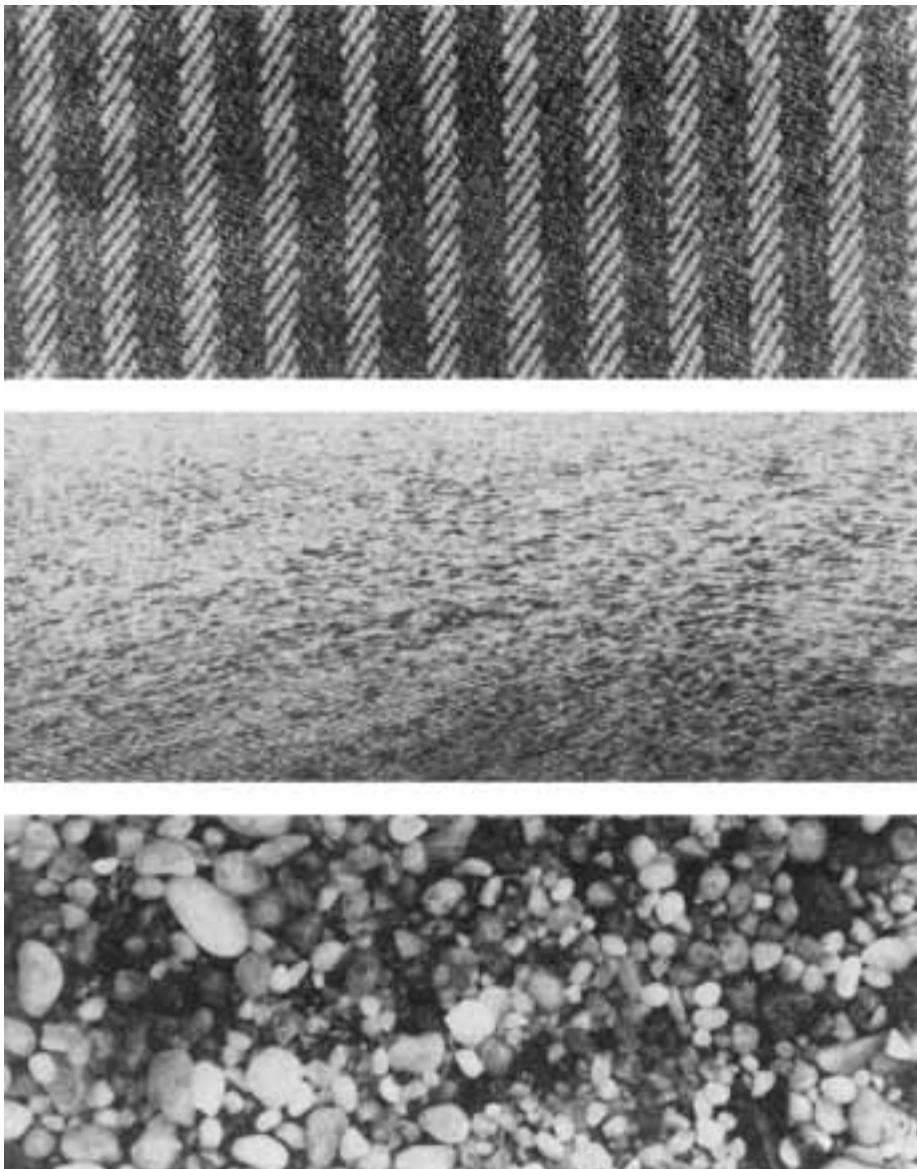
**Figure 3.10** The horizon ratio relation: same height objects at different distances



**Figure 3.11** The horizon ratio relation: different height objects at same distance

One of the most important forms of invariant information in Gibson’s theory is **texture gradient**, although he also discusses gradients of colour, intensity and disparity. There are three main forms of texture gradient relating to the density, perspective and compression of texture elements. The exact nature of a texture element will change from surface to surface (see Figure 3.12); in a carpet the elements are caused by the individual twists of material, on a road they are caused by

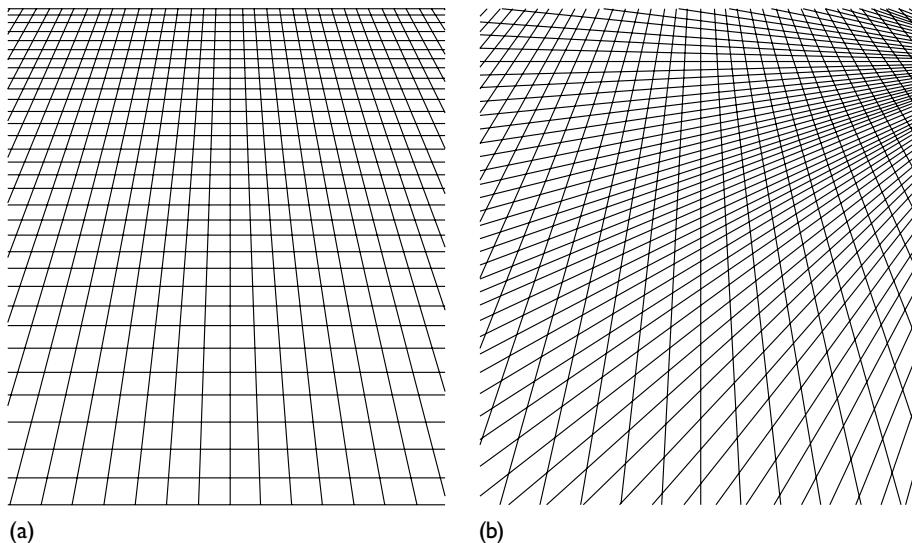
the small stones that make up the surface. In making use of texture gradients, we assume that the texture of the surface is uniform; for example, that the road surface consists of stones of similar size throughout its length. Therefore, any change in the apparent nature of the texture provides us with information regarding the distance, orientation and curvature of the surface.



**Figure 3.12** Examples of texture elements

Source: Gibson, 1979, Figure 2.1

Using texture gradients as a guide, we can tell if a surface is receding because the density of texture elements (number of elements per square metre) will increase with distance. For example, the surface in Figure 3.13(a) appears to recede as the density of texture elements (the individual squares) increases toward the top of the image.



**Figure 3.13** (a) How texture gradient can reveal that a surface is receding; (b) How perspective and compression gradients reveal the shape and orientation of a surface

In a similar fashion, the perspective gradient (the width of individual elements) and the compression gradient (the height of individual elements) can reveal the shape and orientation of a surface. As you can see from Figure 3.13(b), we do not see this surface as flat because the width and height of the individual texture elements changes, making the surface appear to be slanting and curved.

Without texture, considerable ambiguity about shape and orientation can be introduced into the stimulus and this poses a problem for experiments that make use of planar geometric shapes (as you saw with the Necker cube in Activity 3.1). So, texture gradient is a powerful source of invariant information provided by the structure of light within the optic array. It furnishes us with a wealth of information regarding the distance, size and orientation of surfaces in the environment.

### 3.3 Flow in the ambient optic array

What is clear to me now that was not clear before is that structure as such, frozen structure, is a myth, or at least a limiting case. Invariants of structure do not exist except in relation to variants.

*(Gibson, 1979, p.87)*

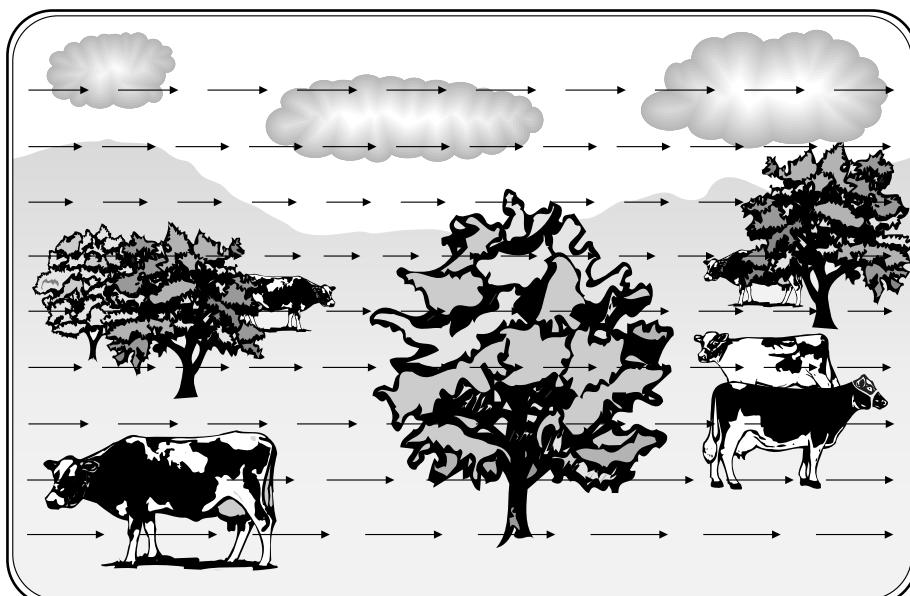
In the above quotation Gibson is highlighting the importance of another intrinsic aspect of perception that is often missing from laboratory stimuli – that of motion. His argument is that invariant information can only be perceived in relation to variant information. To put it another way, in a static view all information is invariant because it never changes. To perceive invariant information, we have to see the environment change over time.

There are two basic forms of movement: motion of the observer and motion of objects within the environment. Motion of the observer tends to produce the greatest degree of movement as the entire optic array is transformed (see Figure 3.9). Gibson

suggested that this transformation provided valuable information about the position and shape of surfaces and objects. For example, information about shape and particularly position is revealed by a phenomenon known as **motion parallax**. The principle of motion parallax is that the further an object is from an observer, the less it will *appear* to move as the observer travels past it. Imagine the driver of a moving inter-city train looking out of their side-window at a herd of cows grazing in a large field next to the line. The cows near the train will *appear* to move past much faster than the cows at the back of the field. Thus, the degree of apparent motion is directly related to the distance of the object from the observer.

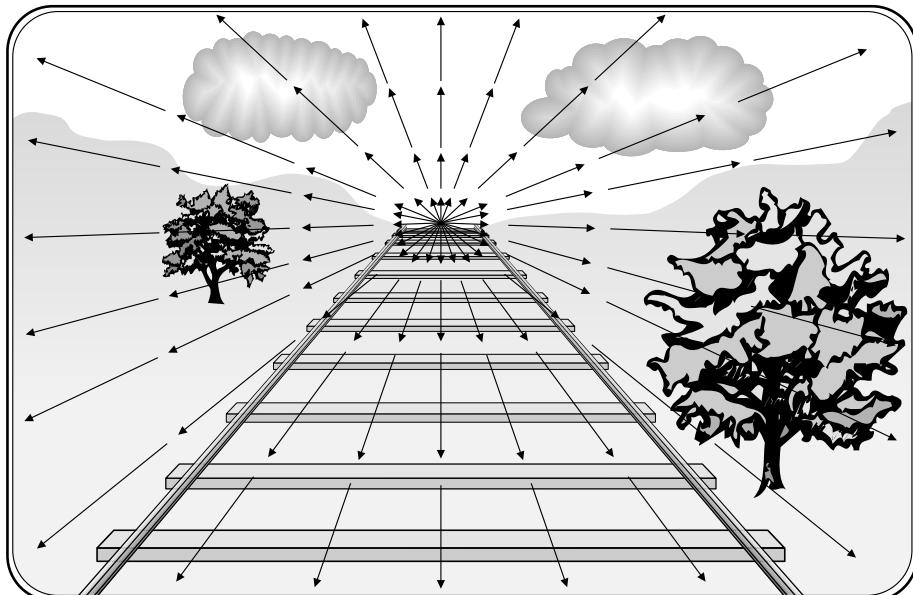
A second means by which observer motion can provide information about the shape and position of objects is through **occlusion**. Imagine the same observer described above travelling past the same field of cows. Their motion will cause the cows nearest to the train to pass in front of, or *occlude*, the cows grazing further away. This allows the observer to deduce that the *occluded* cows (i.e. the ones that become hidden by other cows) are further away than those doing the *occluding*.

Gibson dealt with the motion of the observer through reference to **flow patterns** in the optic array. As our train driver looks at the grazing cows by the side of the track, the entire optic array will appear to flow past from left to right, assuming that the driver looks out of the right-hand window (see Figure 3.14).



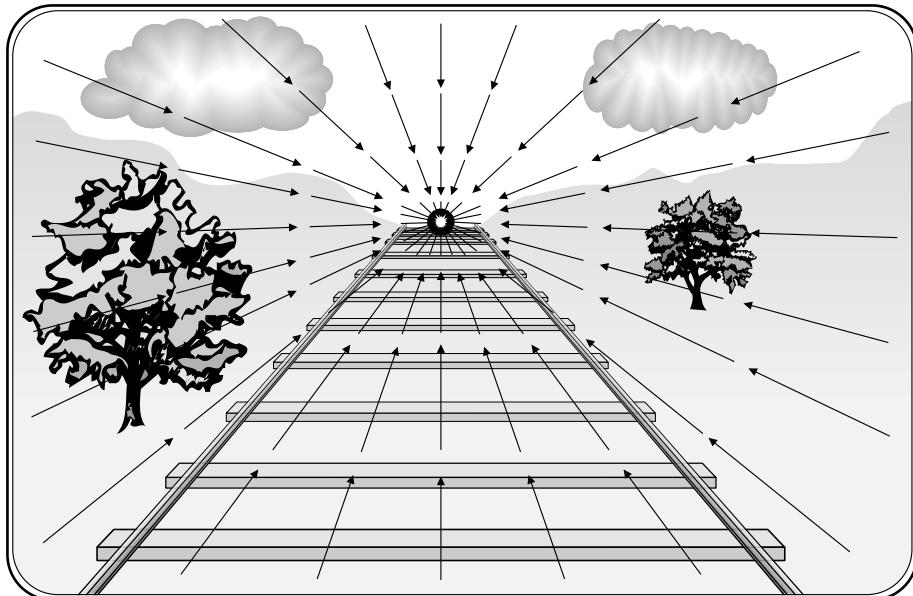
**Figure 3.14** Flow patterns in the optic array parallel to the direction of the observer's motion

When the train driver becomes bored of cow watching and returns their attention to the track in front of the train, the flow patterns in the optic array will change so that the texture elements appear to be radiating from the direction in which the train is travelling (the apparent origin of this radiating flow pattern is known as the **pole**). The texture elements that make up the surfaces in the environment will appear to emerge from the pole, stream toward the observer and then disappear from view (see Figure 3.15).



**Figure 3.15** Flow patterns in the optic array in the direction of the observer's motion

This pattern would be completely reversed if the guard at the rear of the train were to look back toward the direction from which the train had come (see Figure 3.16).



**Figure 3.16** Flow patterns in the optic array in the opposite direction to the observer's motion

Gibson proposed a set of rules that linked flow in the optic array to the movement of the observer through the environment (Gibson, 1979):

- If there is flow in the ambient optic array, the observer is in motion; if there is no flow, the observer is not moving.

- Outflow of the optic array from the pole specifies approach by the observer and inflow to the pole specifies retreat.
- The direction of the pole specifies the direction in which the observer is moving.
- A change in the direction of the pole specifies that the observer is moving in a new direction.

For Gibson, the movement of the observer was a critical part of perception. In fact he deemed it of such importance that he saw the *perceptual system* as not being limited to the eyes and other sense organs but constituting a hierarchy of organs in which the eyes are linked to a head that can turn, which is linked to a body that can move. As Gibson said: ‘perceiving is an act, not a response, an act of attention, not a triggered impression, an achievement, not a reflex’ (Gibson, 1979, p.21).

### 3.4 Affordances and resonance

We began our discussion of Gibson’s theory by stating that he saw information as being directly perceived or ‘picked up’ from the environment. In his later work Gibson (1979) took this idea of information being ‘picked up’ one step further and suggested that the end point of the perceptual process was not a visual description of the surrounding world, but rather that objects directly ‘afforded’ their use.

At its simplest (and least controversial level) the concept of **affordance** builds on earlier research conducted by the Gestalt psychologists, in which the features of objects were seen as providing information as to their use. For instance, the features of a rock would suggest that it could be stood upon, the features of a fallen branch that it could be picked up, and the features of a fruit that it could be eaten.

However, Gibson makes two claims regarding affordances that are rather harder to accept and have proven to be far more controversial. First, he states that affordances act as a bridge between perception and action and do not require the intervention of any cognitive processes. Just as the nature of the environment can be directly ‘picked up’ from the structure of the optic array, the observer can interact with surfaces and objects in the environment directly through affordance.

Second, Gibson saw no role for memory in perception, as the observer does not have to consult their prior experience in order to be able to interact with the world around them. Instead he states that the perceptual system **resonates** to invariant information in the optic array. Although the definition of ‘resonates’ and the identity of what is doing the resonating is left very vague by Gibson, the point is that ‘global’ information about the optic array (in the form of invariant information) is dealt with by the perceptual system without the need to analyse more ‘local’ information such as lines and edges.

These assertions may seem unreasonable to you, as they have done to other researchers. If we are studying psychology, then surely the cognitive processes that allow us to perceive must be one focus of our attention. In addition, if when perceiving the world we do not make use of our prior experiences, how will we ever learn from our mistakes? In the next two sections we shall turn to theories that attempt to deal with these issues and to explain exactly how the brain makes sense of the world around us.

However, even if Gibson's theory does not enlighten us as to the nature of the cognitive processes that are involved in perception, his theory has been extremely influential, and researchers in perception still need to bear in mind his criticisms of the laboratory approach which makes use of artificial stimuli:

Experiments using dynamic naturalistic stimuli can now be conducted, virtual scenes can be constructed, and images of brain activity while viewing these can be captured in a way that would have been difficult to envisage a century ago. However, the simulated lure of the screen (or even a pair of screens) should not blind experimenters and theorists to the differences that exist between the virtual and the real.

(Wade and Bruce, 2001, p.105)

## Summary of Section 3

---

- Gibson developed an ecological approach to perception and placed great emphasis on the way in which real objects and surfaces structure light – he termed this the ambient optic array.
- He suggested that invariant information (such as texture gradient) could be 'picked up' from the optic array to provide cues as to the position, orientation and shape of surfaces.
- Invariant information could also be revealed by motion, which produces variants such as flow patterns in the optic array.
- The importance of real surfaces and of motion led Gibson to suggest that perception could not be studied using artificial stimuli in a laboratory setting.
- Gibson did not see perception as a product of complex cognitive analysis, but suggested that objects could 'afford' their use directly.
- Interaction with the environment is at the heart of Gibson's theory; action is seen as the 'goal' of perception.

## 4 Marr's theory of perception

... the detection of physical invariants, like image surfaces, is exactly and precisely an information-processing problem, in modern terminology. And second, he (Gibson) vastly underrated the sheer difficulty of such detection ... Detecting physical invariants is just as difficult as Gibson feared, but nevertheless we can do it. And the only way to understand how is to treat it as an information-processing problem.

(Marr; 1982, p.30)

As we stated previously, one criticism that has been levelled at Gibson's approach is that it does not explain in sufficient detail *how* information is picked up from the

environment. To address this problem, a theory was needed that attempted to explain exactly how the brain was able to take the information sensed by the eyes and turn it into an accurate, internal representation of the surrounding world. Such a theory was proposed by David Marr (1982).

Before we look at Marr's theory, it is worth pointing out some of the similarities and differences between the approaches taken by Marr and Gibson. Like Gibson, Marr's theory suggests that the information from the senses is sufficient to allow perception to occur. However, unlike Gibson, Marr adopted an information-processing approach in which the processes responsible for analysing the retinal image were central. Marr's theory is therefore strongly 'bottom-up', in that it sees the retinal image as the starting point of perception and explores how this image might be analysed in order to produce a description of the environment. This meant that, unlike Gibson who saw action as the end point of perception, Marr concentrated on the perceptual processes involved in object recognition.

Marr saw the analysis of the retinal image as occurring in four distinct stages, with each stage taking the output of the previous one and performing a new set of analyses on it. The four stages were:

- 1 *Grey level description* – the intensity of light is measured at each point in the retinal image.
- 2 *Primal sketch* – first, in the raw primal sketch, areas that could potentially correspond to the edges and texture of objects are identified. Then, in the full primal sketch, these areas are used to generate a description of the outline of any objects in view.
- 3 *2½D sketch* – at this stage a description is formed of how the surfaces in view relate to one another and to the observer.
- 4 *3D object-centred description* – at this stage object descriptions are produced that allow the object to be recognized from any angle (i.e. independent of the viewpoint of the observer).

More generally, Marr concentrated his work at the computational theory and algorithmic levels of analysis (see Chapter 1) and had little to say about the neural hardware that might be involved. One reason for this is that he developed his theory largely by designing computer-based models and algorithms that could perform the requisite analyses.

## 4.1 The grey level description

One way of describing the first stage in Marr's theory is to say that it gets rid of colour information. This is not because Marr thought that colour was unimportant in perception. Rather, he thought that colour information was processed by a distinct **module** and need not be involved in obtaining descriptions of the shape of objects and the layout of the environment. In fact, the modular nature of perception was a fundamental part of Marr's theory:

Computer scientists call the separate pieces of a process its modules, and the idea that a large computation can be split up and implemented as a collection of parts that are as nearly independent of one another as the

overall task allows, is so important that I was moved to elevate it to a principle; the principle of modular design.

(Marr, 1982, p.102)

This meant that the perception of colour could be handled by one ‘module’ and the perception of shape by another.

The first stage in Marr’s theory acts to produce a description containing the intensity (i.e. the brightness) of light at all points of the retina. A description composed solely of intensity information is referred to as ‘greyscale’, as, without the information provided by analysing the wavelength of light, it will consist of nothing but different tones of grey. If you turn down the colour on your TV, the resulting picture will be a greyscale image – although we call it ‘black and white’, it actually consists of many shades of grey.

Without going into too much detail, it is possible to derive the intensity of the light striking each part of the retina, because as light strikes a cell in the retina, the voltage across the cell membrane changes and the size of this change (or **depolarization**) corresponds to the intensity of the light. Therefore, a greyscale (or grey level) description is produced by the pattern of depolarization on the retina. In other words it is possible to derive the greyscale description simply by analysing the outputs of the receptor cells in the retina.

## 4.2 The primal sketch

The next part in Marr’s theory, the generation of the primal sketch, occurs in two stages. The first stage consists of forming a raw primal sketch from the grey level description by identifying patterns of changing intensity.

### ACTIVITY 3.7

---

Find a wooden table or chair and place it where it is both well-illuminated and against a light background. Describe how the intensity of the light reflected from the table/chair changes across its surface and in comparison with the background.

### COMMENT

You should be able to see that the edges of the table/chair are marked by a quite large, sharp change in the intensity of the reflected light caused by the object in question being darker than the background. In addition, there are smaller changes in intensity caused by the individual parts of the table/chair and by the texture of the wood. You may also have noticed other changes in the intensity of the reflected light that did not correspond to the edge of the object, its parts or texture.

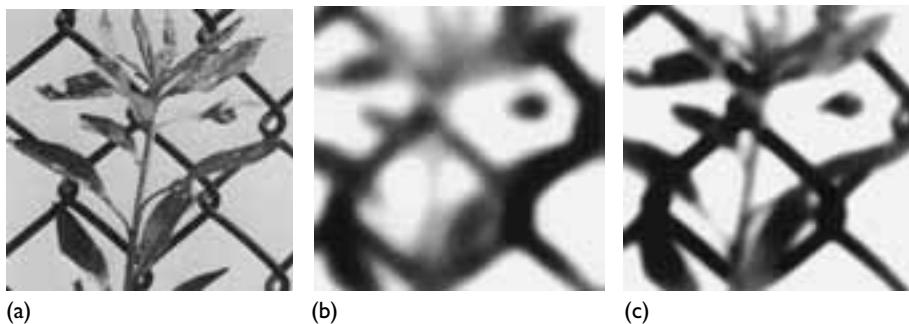
---

It is possible to group changes in the intensity of the reflected light into three categories:

- Relatively large changes in intensity produced by the edge of an object.
- Smaller changes in intensity caused by the parts and texture of an object.
- Still smaller changes in intensity due to random fluctuations in the light reflected.

Marr and Hildreth (1980) proposed an algorithm that could be used to determine which intensity changes corresponded to the edges of objects, meaning that changes in intensity due to random fluctuations could be discarded. The algorithm made use of a technique called **Gaussian blurring**, which involves averaging the intensity values in circular regions of the greyscale description. The values at the centre of the circle are weighted more than those at the edges in a way identical to a normal (or Gaussian) distribution.

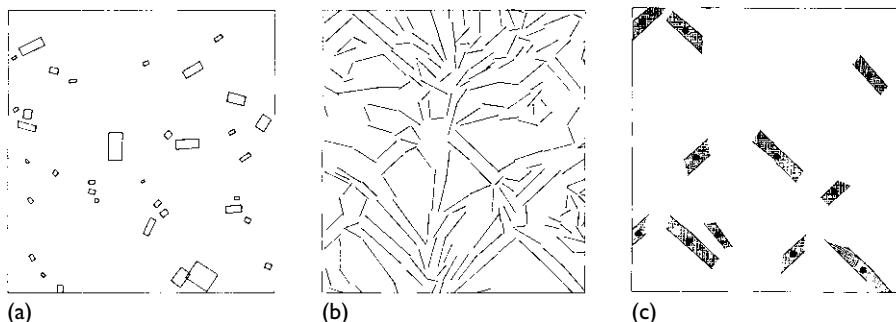
By changing the size of the circle in which intensity values are averaged, it is possible to produce a range of images blurred to different degrees. Figure 3.17 shows images that have been produced in this manner. The original (i.e. unblurred) image is shown in (a). As you can see, using a wider circle (b) produces a more blurred image than using a narrower circle (c).



**Figure 3.17** Examples of Gaussian blurred images

Source: Marr and Hildreth, 1980, p.190

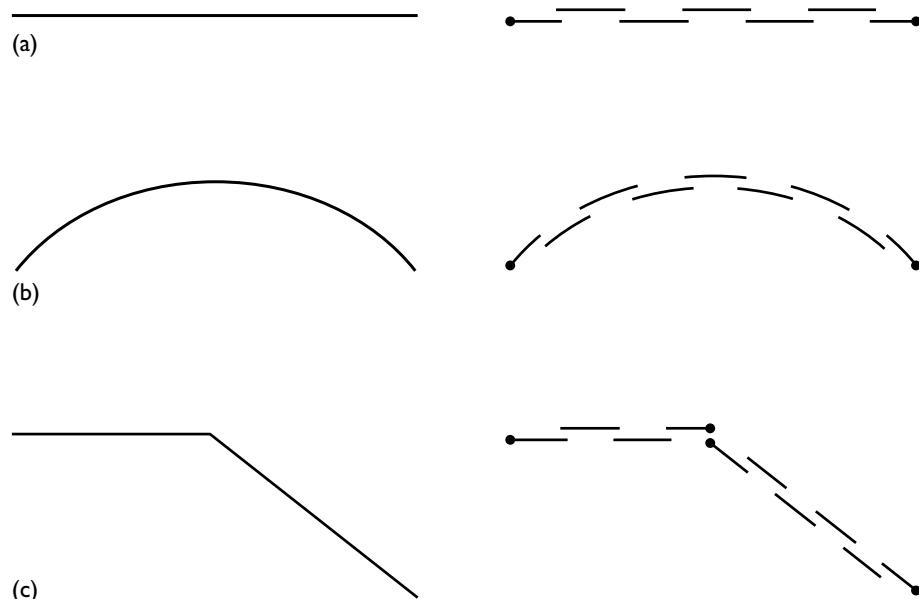
Marr and Hildreth's algorithm works by comparing images that have been blurred to different degrees. If an intensity change is visible at two or more adjacent levels of blurring, then it is assumed that it cannot correspond to a random fluctuation and must relate to the edge of an object. Although this algorithm was implemented by Marr and Hildreth on a computer, there is evidence that retinal processing delivers descriptions that have been blurred to different degrees.



**Figure 3.18** Primitives used in the raw primal sketch: (a) blobs, (b) edge-segments and (c) bars

Source: Marr, 1982, Figure 2.21, p.72

By analysing the changes in intensity values in the blurred images, it is possible to form a symbolic representation consisting of four **primitives** corresponding to four types of intensity change. Marr referred to these primitives as ‘edge-segments’, ‘bars’, ‘terminations’ and ‘blobs’. An edge-segment represented a sudden change in intensity; a bar represented two parallel edge-segments; a termination represented a sudden discontinuity; and a blob corresponded to a small, enclosed area bounded by changes in intensity. In Figure 3.18, you can see how the image shown in Figure 3.17(a) would be represented using three of these primitives, whilst Figure 3.19 shows how three simple lines would be represented in the raw primal sketch.



**Figure 3.19** Representation of three simple lines in the raw primal sketch: ‘The raw primal sketch represents a straight line as a termination, several oriented segments, and a second termination (a). If the line is replaced by a smooth curve, the orientations of the inner segments will gradually change (b). If the line changes its orientation suddenly in the middle (c), its representation will include an explicit pointer to this discontinuity. Thus in this representation, smoothness and continuity are assumed to hold unless explicitly negated by an assertion’ (Marr, 1982, p.74)

Source: Marr, 1982, Figure 2.22, p.74

As you can see from Figure 3.19, although the raw primal sketch contains a lot of information about details in the image, it does not contain explicit information about the global structure of the objects in view. The next step is therefore to transform the raw primal sketch into a description, known as the **full primal sketch**, which contains information about how the image is organized, particularly about the location, shape, texture and internal parts of any objects that are in view.

Basically, the idea is that **place tokens** are assigned to areas of the raw primal sketch based on the **grouping** of the edge-segments, bars, terminations and blobs. If these place tokens then form a group themselves, they can be aggregated together to form a new, higher-order place token.

Imagine looking at a tiger. The raw primal sketch would contain information about the edge of the tiger's body, but also about the edges and pattern of its stripes and the texture of its hair. In the full primal sketch, place tokens will be produced by the grouping of the individual hairs into each of the stripes. The place tokens for each stripe would then also be grouped (because they run in a consistent vertical pattern along the tiger) into a higher-order place token, meaning that there will be at least two levels of place tokens making up the tiger.

Various mechanisms exist for grouping the raw primal sketch components into place tokens and for grouping place tokens together. These include **clustering**, in which tokens that are close to one another are grouped in a way very similar to the Gestalt principle of proximity, and **curvilinear aggregation**, in which tokens with related alignments are grouped in a similar fashion to the Gestalt principle of good continuation.

As we saw in Section 2, perceptual grouping is a robust, long-established and powerful effect. Marr saw algorithms expressing laws such as those formulated by the Gestalt approach as being responsible for turning the ambiguous raw primal sketch into the full primal sketch in which the organization of objects and surfaces was specified.

### 4.3 The 2½D sketch

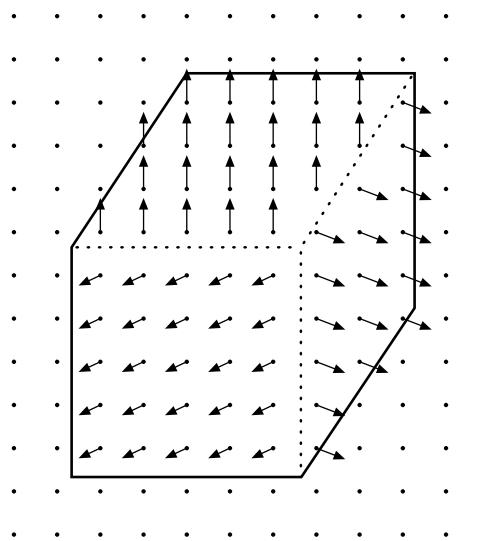
In Marr's theory, the goal of early visual processing is the production of a description of the environment in which the layout of surfaces and objects is specified in relation to the particular view that the observer has at that time. Up until now we have been looking at how the shape of objects and surfaces can be recovered from the retinal image. However, in order to specify the layout of surfaces, we need to now include other information, specifically cues that tell us how far away each surface is.

Marr's modular approach to perception means that while the full primal sketch is being produced, other visual information is being analysed simultaneously. Much of this has to do with establishing depth relations, the distance between a surface and the observer and also how far objects extend. We saw in Section 3 that motion cues and cues from texture can be used to specify the distance to an object, and it is also possible to make use of the disparity in the retinal images of the two eyes (known as **stereopsis**), and shading cues that are represented in the primal sketch.

Marr proposed that the information from all these 'modules' was combined together to produce the 2½D sketch. It is called the 2½D sketch, rather than the 3D sketch, because the specification of the position and depth of surfaces and objects is done in relation to the observer. Thus, the description of an object will be **viewer-centred** and will not contain any information about the object that is not present in the retinal image. How the viewer-centred 2½D sketch is turned into a fully 3D, **object-centred** description is one of the topics dealt with in the next chapter.

Marr saw the 2½D sketch as consisting of a series of primitives that contained vectors (a line depicting both size and direction) showing the orientation of each surface. A vector can be seen as a needle, in which the direction the needle is pointing tells us in which direction the surface is facing, and its length tells us by how much the surface is slanted in relation to the observer. A cube would therefore be

represented like the one shown in Figure 3.20. In addition to the information shown in Figure 3.20, Marr suggested that each vector (or needle) would have a number associated with it that indicated the distance from the observer.



**Figure 3.20** A 2½D sketch of a cube

Source: Marr, 1982, Figure 4.2, p.278

The 2½D sketch therefore provides an unambiguous description of the size, shape, location, orientation and distance of all the surfaces currently in view, in relation to the observer.

#### 4.4 Evaluating Marr's approach

Marr's theory was the catalyst for a great deal of computational and psychological research. Some of this research has reported findings consistent with the mechanisms proposed by Marr, whilst some has found that Marr's theory does not offer a good explanation for the results obtained. We will not attempt to review every single study here, but instead describe a few studies that have tested elements of Marr's theory.

Marr and Hildreth (1980) attempted to test their idea that the raw primal sketch was formed by searching for changes in intensity values in adjacent levels of blurring, by implementing this algorithm in a computer program. They found that when applied to images of everyday scenes the algorithm was reasonably successful in locating the edges of objects. However, as with all computer-simulation research, it is important to remember that, just because a specific program yields the expected results, it does not necessarily follow that this is what is happening in the human perceptual system.

It seems as if Marr's approach to the formation of the full primal sketch was flawed in that it was limited to grouping strategies based on the 2D properties of an image. Enns and Rensick (1990) showed that participants could easily determine which one of a series of figures consisting of blocks was the odd-one-out, even

though the only difference between the figures was their orientation in three dimensions. Thus, some grouping strategies must make use of 3D information.

One area in which Marr's theory does seem to fit the results of experimentation is in the integration of depth cues in the 2½D sketch. This phenomenon has been studied in experiments that have attempted to isolate certain forms of depth cue and then determine how they interact. For example, Young *et al.* (1993) looked at how motion cues interacted with texture cues. They concluded that the perceptual system does process these cues separately, and will also make selective use of them depending on how 'noisy' they are. In other words, in forming the 2½D sketch, the perceptual system does seem to integrate different modules of depth information, but will also place more emphasis on those modules that are particularly useful for processing the current image.

As well as the success of the specific processes suggested by Marr, it is also possible to evaluate his theory according to broader concepts. As we shall see in Section 6, there is evidence that there are two visual pathways in the brain that appear to process separately 'what' information and 'where' information. It seems that there exist different perceptual processes according to whether the goal of perception is action or object recognition. Although Marr's theory is a modular approach, so that different types of visual information are processed separately, it did not predict the separation of visual pathways into action and object recognition and indeed it is hard to incorporate this into the theory (Wade and Bruce, 2001). However, although the precise nature of the processes suggested by Marr may not map exactly onto those actually used by the brain to perceive the world, the impact of Marr's theory should not be underestimated: 'Thus it is not the details of Marr's theory which have so far stood the test of time, but the approach itself' (Wade and Bruce, 2001, p.97).

## Summary of Section 4

---

- Marr proposed a theory of vision that was based on bottom-up processing of information.
- His approach was to see perception as being composed of a series of stages, with each stage generating an increasingly sophisticated description.
- Marr saw the end point of the perceptual process as object recognition rather than action.
- The first stage involves producing a grey level description based on the activation of retinal cells.
- This description is analysed by blurring it to different degrees. Changes in intensity value that are present in two or more adjacent levels of blurring are assumed to correspond to the 'edge' of an object (or part of an object).
- The raw primal sketch is generated by assigning one of four primitives (edge-segment, bar, termination or blob) to each change in intensity values.

- The full primal sketch is generated by using perceptual organizational principles such as clustering and similarity to group these primitives together and assign each group a place token.
- Information from different modules (such as stereopsis and motion) are combined with the full primal sketch to produce the 2½D sketch. This contains primitives consisting of vectors that reveal the distance and orientation, in relation to the observer, of the visible surfaces.

## 5 Constructivist approaches to perception

The previous sections of this chapter should have given you some idea of how we can see and interpret sensory information. The emphasis so far has been on ‘bottom-up’ processes. As discussed previously, there is also information flowing ‘top-down’ from stored knowledge. This makes intuitive sense. To be able to perceive something as ‘a bus’, you need to access stored knowledge concerning what the features of a bus actually are (big object with wheels etc.).

Thus, what you see a stimulus *as* depends on what you know. This notion that perceiving something involves using stored knowledge as well as information coming in from the senses is embodied in an approach referred to as the **constructivist approach**. The approach is described as ‘constructivist’ because it is based on the idea that the sensory information that forms the basis of perception is, as we have already suggested, incomplete. It is necessary to *build* (or ‘construct’) our perception of the world from *incomplete* information. To do this we use what we already *know* about the world to interpret the incomplete sensory information coming in, and to ‘make sense’ of it. Thus stored knowledge is used to aid in the recognition of objects.

### ACTIVITY 3.8

---

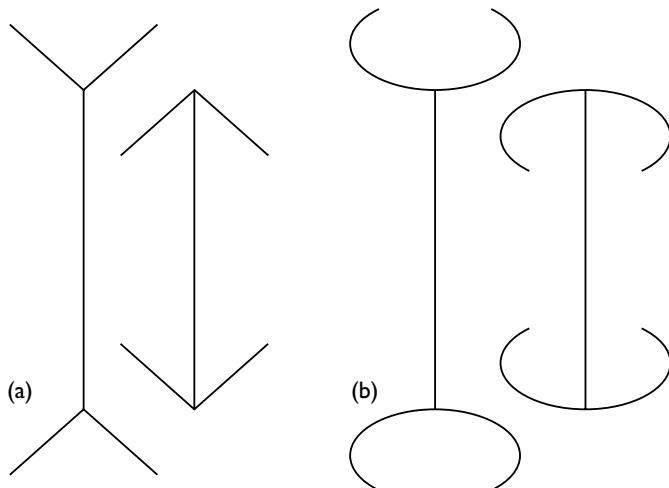
Look back at Activity 3.1. Can you explain any of the visual illusions in terms of what you now know about the bottom-up approach to perception?

### COMMENT

Gibson would tell us that the Necker cube is a geometric figure that contains none of the information (particularly texture gradients) that we would usually use when perceiving an object. Marr’s theory can help us to explain Kanizsa’s illusory square, as the four areas of intensity change corresponding to the missing parts of the circles would be grouped together to form a square.

But what about the Müller-Lyer illusion? There are a number of alternative explanations for this illusion, one of which is that we group each vertical line with its set of arrowheads to form a single object. This of course results in the object with the inward-pointing arrowheads being larger than the one with the outward-pointing arrowheads; basically, due to perceptual grouping we cannot separate the vertical line from the overall size of the object. However, as the Müller-Lyer illusion is reduced if

the straight arrowheads are replaced with curved lines (see Figure 3.21), it could be that we also need to look at an explanation based on top-down perception.



**Figure 3.21** The original Müller-Lyer illusion (a), and with curved arrowheads (b)

As we saw in Activity 3.1, another explanation of the Müller-Lyer illusion is that we make use of top-down information and see the outward-pointing arrowheads as an indication that the vertical line is nearer to us than the line with the inward-pointing arrowheads.

---

Two of the foremost proponents of the constructivist approach are Irvin Rock (1977, 1983, 1997) and Richard Gregory (1980). Gregory suggested that individuals attempt to recognize objects by generating a series of **perceptual hypotheses** about what that object might be. Gregory conceptualized this process as being akin to how a scientist might investigate a problem by generating a series of hypotheses and accepting the one that is best supported by the data (in perception, ‘data’ would be the information flowing ‘up’ from the senses).

We are forced to generate hypotheses, according to Gregory’s argument, because the sensory data are incomplete. If we had perfect and comprehensive sensory data we would have no need of hypotheses as we would *know* what we perceived. Stored knowledge is assumed to be central to the generation of perceptual hypotheses as it allows us to fill in the gaps in our sensory input. The influence of stored knowledge in guiding perceptual hypotheses can be demonstrated by the use of impoverished figures such as the one in Figure 3.22 (Street, 1931).

At first glance this picture may be difficult to perceive as anything other than a series of blobs. So the resulting hypothesis might be that it is just, ‘a load of blobs.’ If however, you are told that it is a picture of an ocean liner (coming towards you, viewed from water level) then the picture may immediately resolve into an image of an ocean liner. The sensory information has not changed, but what you know about it has, allowing you to generate a reasonable hypothesis of what the figure represents. Similarly, in the example, used in Activity 3.3, of trying to



**Figure 3.22** An example of an impoverished figure

Source: Street, 1931

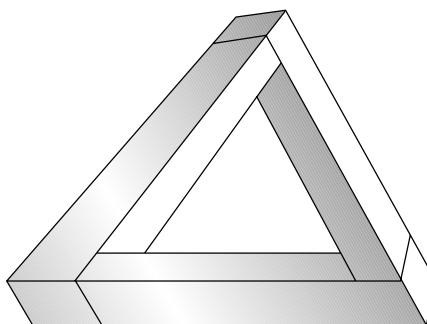
identify an object by touch alone, if you are given some clues about the function of the object (i.e. your knowledge related to the object is increased), it is likely to be easier to identify it.

The use of knowledge to guide our perceptual hypotheses may not always lead to a ‘correct’ perception. There are some stimuli with which we are so familiar (such as faces) that there can be a strong bias towards accepting a particular perceptual hypothesis, resulting in a ‘false’ perception. For instance, look at the faces in Figure 3.23.



**Figure 3.23** The mask of Hor

This is the mask of Hor, an Egyptian mummy. The first view is the mask from the front and the second two are of the back of the mask. Although the face viewed from the back is ‘hollow’ it still appears perceptually as a normal face. Our knowledge of how a face is supposed to look is (according to Gregory, 1980) so strong that we cannot accept the hypothesis that a face could be ‘hollow.’ This effect is interesting in that it provides an example of a perceptual hypothesis conflicting with what Gregory terms ‘high-level’ knowledge. You *know* at a conceptual level that the mask is hollow, yet you still *perceive* it as a ‘normal’ face. This, as Gregory suggests, represents a tendency to go with the most *likely* hypothesis. The Penrose triangle (Penrose and Penrose, 1958) in Figure 3.24 demonstrates a similar point.



**Figure 3.24** An impossible triangle

Source: Penrose and Penrose, 1958

It would be impossible to construct the object in Figure 3.24 so that the three sides of the triangle were joined. At one level, we ‘know’ that this must be true. Yet whichever corner of the triangle we attend to suggests a particular 3D interpretation. Our interpretation of the figure changes as our eyes (or just our attention) jumps from corner to corner. These data-supported interpretations, or hypotheses, tend to overwhelm the conceptual knowledge that we are viewing a flat pattern.

Although the constructivist approach in general, and Gregory’s theories in particular, provide an attractive explanatory framework for perception, there are areas of the theory (as there were with Gibson’s approach) that are left rather vague. For instance, how do we actually generate hypotheses and how do we know when to stop and decide which is the ‘right’ one? Why does knowledge sometimes but not always help perception? How can we ‘know’ something is wrong, and yet still perceive it as wrong (as with the hollow face)? Although these are difficult questions to answer, progress is being made in explaining how human perception may be based, at least in part, on constructivist principles; some of this work will be discussed below.

Thus, there appears to be evidence that perceptions of the outside world can be ‘constructed’ using information flowing ‘up’ from the senses combined with knowledge flowing ‘down’. However, this seems to be in direct contrast to the theories of Gibson and Marr discussed earlier which suggest that there is no need to use stored knowledge to interpret the information flowing in from the senses. Indeed, the impossible triangle above shows that we do not always make use of knowledge that may be relevant and available. So, just how important is knowledge to the process of perception, and is there any way in which we can reconcile theories of perception that see knowledge as being essential with those that see it as unnecessary? The following section considers how these different theories may be reconciled through consideration of the way in which the brain processes sensory information.

## Summary of Section 5

---

- What you see a stimulus *as* depends on what you know. This means that perception must involve top-down processing.
- The constructivist approach to perception is based on the idea that sensory data is often incomplete, so a description can only be constructed by making use of stored knowledge.
- Gregory suggested that sensory data are incomplete and we perceive by generating a series of perceptual hypotheses about what an object might be.
- The use of stored information can lead to perceptual hypotheses that are inaccurate, which is why we may be fooled by some visual illusions.

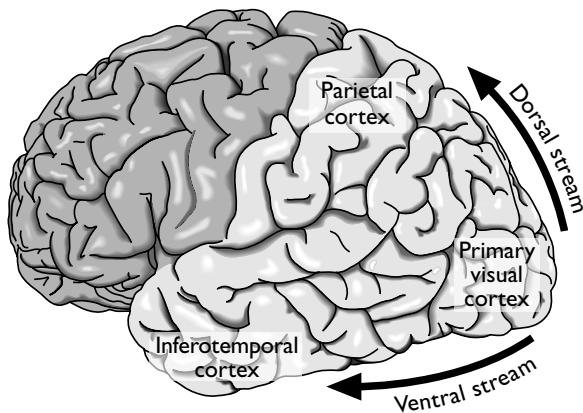
## 6 The physiology of the human visual system

There appear to be at least two (and maybe more) partially distinct streams of information flowing back from the retina (via the optic nerve) into the brain (e.g. Shapley, 1995). The characteristics of these streams and their relation to the theories of perception already described is the topic of this section. It should be emphasized that the distinction between the two streams is fairly loose. There is overlap in the types of information that the streams carry and there are numerous interconnections between them, but they may conveniently be conceptualized as distinct. The following subsections trace these streams of information from the retina to the brain.

### 6.1 From the eye to brain

You may remember from Section 1.2 that there are two types of light-sensitive cells in the retina, called rods and cones. Both rods and cones are connected to what are termed retinal ganglion cells that essentially connect the retina to the brain. Ganglion cell axons leave the eye via the ‘blind spot’ (the concentration of blood vessels and nerve axons here means that there is no room for any receptors, hence this region is ‘blind’). These cells then project (send connections) to an area termed the lateral geniculate nucleus (LGN), and from there to the area of the brain known as the ‘primary visual cortex’ (also known as V1). Even at the level of retinal ganglion cells, there is evidence of two distinct streams or ‘pathways,’ referred to as the parvocellular pathway, and the magnocellular pathway (e.g. Shapley, 1995). These names derive from the relative sizes of the cells in the two pathways, larger cells in the magnocellular pathway, and smaller cells in the parvocellular one. This distinction is maintained up to, and within, the primary visual cortex, although there are interconnections between the two pathways.

Information travelling onward from the primary visual cortex is still maintained in two distinct streams (see Figure 3.25). One stream, leading to the inferotemporal cortex, is termed the **ventral stream**, and the other, leading to the parietal cortex, is known as the **dorsal stream** (these were described briefly in Chapter 1, Section 5.1).



**Figure 3.25** The dorsal and ventral streams

## 6.2 The dorsal and ventral streams

The *ventral* stream projects to regions of the brain that appear to be involved in pattern discrimination and object recognition, whilst the *dorsal* stream projects to areas of the brain that appear to be specialized for the analysis of information about the position and movement of objects. Schneider (1967, 1969) carried out work with hamsters which suggested that there were two distinct parts of the visual system, one system concerned with making pattern discriminations, the other involved with orientation in space. Schneider suggested that one system is concerned with the question, ‘What is it?’, whereas the other system is concerned with the question, ‘Where is it?’. This, and later, work (Ungerleider and Mishkin, 1982) led to the ventral pathway being labelled a ‘what’ system, and the dorsal pathway a ‘where’ system.

Although the two streams appear to be specialized for processing different kinds of information, there is ample evidence of a huge degree of interconnection between the systems at all levels. Also, the streams appear to converge in the prefrontal cortex (Rao *et al.*, 1997), although there is still some evidence that the dorsal–ventral distinction is maintained (Courtney *et al.*, 1996). It has been suggested that it is in the prefrontal cortex that meaning is associated with the information carried by the two streams.

Although describing the two streams as ‘what’ and ‘where’ is convenient, there is a large body of work that suggests that the distinction is not quite that straightforward. For instance, Milner and Goodale (1995) report a number of studies with a patient, DF, who suffered severe carbon monoxide poisoning that appeared to prevent her using her ventral system for analysing sensory input. She could not recognize faces or objects, or even make simple visual discriminations such as between a triangle and a circle. She could draw objects from memory but not recognize them once she had drawn them. DF did, however, appear to have an intact dorsal stream. Although unable to tell if two discs were of the same or different widths (or even *indicate* the widths by adjusting the distance between her fingers), if she was asked to pick the discs up then the distance between her index finger and thumb as she went to pick them up was highly correlated with the width of the discs.

In other words, she did *not* have size information available to conscious perception (via the ventral stream), but it *was* available to guide *action* (via the dorsal stream).

Norman (2002), following on from similar suggestions by Bridgeman (1992) and Neisser (1994), has drawn on the ongoing debate concerning the characteristics of the dorsal and ventral streams and suggested a dual-process approach. In this approach, the two streams are seen as acting synergistically so that the dorsal stream is largely concerned with perception for action and the ventral stream essentially concerned with perception for recognition. The dual-process approach is supported by some of the characteristics of the two streams (Norman, 2001, 2002):

- 1 There appears to be evidence (Goodale and Milner, 1992; Ungerleider and Mishkin, 1982) to suggest that the ventral stream is primarily concerned with recognition whilst the dorsal stream drives visually guided behaviour (pointing, grasping, etc.).
- 2 The ventral system is generally better at processing fine detail (Baizer *et al.*, 1991) whereas the dorsal system is better at processing motion (Logothesis, 1994).
- 3 The studies on patient DF (Milner and Goodale, 1995) suggest that the ventral system is knowledge-based and uses stored representations to recognize objects, whilst the dorsal system appears to have only very short-term storage available (Bridgeman *et al.*, 1997; Creem and Proffitt, 1998).
- 4 The dorsal system receives information faster than the ventral system (Bullier and Nowak, 1995).
- 5 A limited amount of psychophysical evidence suggests that we are much more conscious of ventral than of dorsal stream functioning (Ho, 1998).
- 6 It has been suggested (Goodale and Milner, 1992; Milner and Goodale, 1995) that the ventral system recognizes objects, and is thus object-centred. The dorsal system is presumed to be used more in driving some action in relation to an object and thus uses a viewer-centred frame of reference (this distinction arises again in the next chapter).

### 6.3 The relationship between visual pathways and theories of perception

We have already seen that Gibson's approach to perception concentrated more on perception for the purposes of action, whilst Marr's theory was principally concerned with object recognition. In addition, the constructivist approach is also more concerned with perception for recognition than perception for action, as it concentrates on how we may use existing knowledge to work out what an object might be. Although these approaches have their differences, it is undoubtedly the case that we need to both recognize objects and perform actions in order to interact with the environment. It could be then, that the type of perception discussed by Gibson is principally subserved by the dorsal system, whilst the ventral system is the basis for the recognition approach favoured by Marr and the constructivists.

For example, Gibson's notion of 'affordance' emphasizes that we might need to detect what things are *for* rather than what they actually *are*. That is, affordances are linked to actions ('lifting' or 'eating', for example). The dorsal system appears to be

ideally suited to providing the sort of information we need to act in the environment. In addition, if a system is to be used to drive action, it really needs to be fast, as the dorsal stream seems to be.

The earlier discussion of Gibson's ecological approach also stated that Gibson saw no need for memory in perception. Certainly, one of the characteristics of the dorsal stream is that it appears to have no more than a very short 'memory' (at least for representations of objects). Thus, there appear to be some grounds for suggesting that the dorsal stream is Gibsonian in operation.

In contrast, the ventral stream appears to be ideally suited to the role of recognizing objects. It is specialized in analysing the sort of fine detail that Marr saw as essential to discriminating between objects, and it also seems able to draw on our existing knowledge (top-down information) to assist in identifying them. In addition, it is slower than the dorsal stream; but then recognizing what an object may be is not necessarily an immediate priority. For example, knowing *that* an object is moving toward you quickly is initially more important than knowing *what* it is.

#### **6.4 A dual-process approach?**

Norman's proposal discussed above does provide an attractive way of reconciling two of the classic approaches to visual perception. There is perhaps a danger, however, in trying to 'shoehorn' what is known about the dorsal and ventral streams into the framework provided by previous theories. Given that both the constructivist and Gibsonian theories are rather vague on how the processes they describe could be implemented, it is questionable how useful they are as a theoretical framework in which to interpret the workings of the dorsal and ventral streams. Attempting to explain the streams in the light of the previous theories does tend to emphasize the way in which they work separately rather than the way in which they work together. Undoubtedly, the two streams *can* operate independently (as demonstrated by the case of DF discussed earlier), but this is rather like saying that you can take the steering column out of a car and both the car and the steering wheel will still function to some degree! In fact, Norman (2002) describes the two streams as synergistic and interconnected, rather than independent.

Binsted and Carlton (2002), in a commentary on the proposal put forward by Norman, provide an illustration of the interaction between the dorsal and ventral streams using the example of skill acquisition. Previous work (Fitts, 1964) suggests that the early stages of learning a skill (such as driving) are characterized by cognitive processes of the sort associated with the ventral stream, whereas once the task is well practised it is characterized by learned motor actions of the sort associated with the dorsal stream.

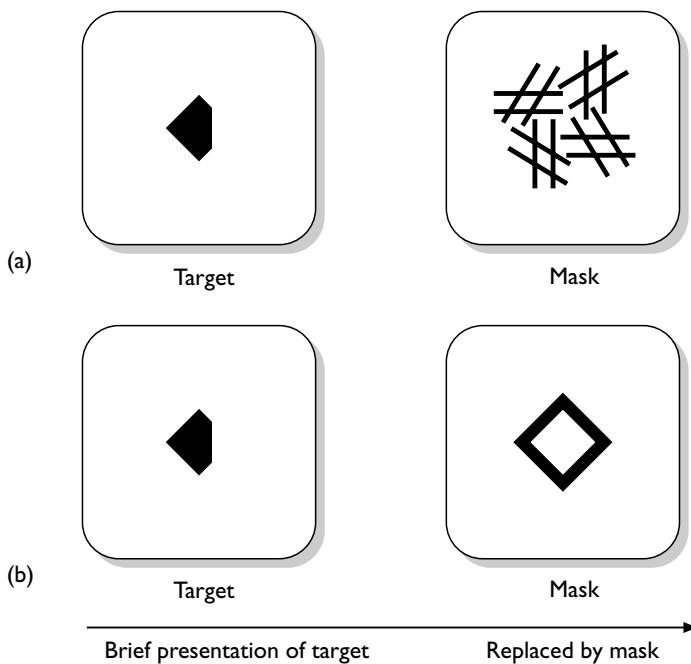
The question is, if these two streams function in such different ways, how is learning transferred from one to the other? It is possible, of course, that learning occurs in both streams at the same time and that whichever is most effective 'leads' in performance of the task, but this still implies a high degree of interaction between them and a blurring of the boundaries between their functions. The issue (which is as yet unresolved) then becomes whether the two streams interact to such an extent that it is meaningless to consider them to be functionally separate and representative of different theoretical approaches to visual processing (as Norman suggests). Thus,

rather than questioning whether both Gibsonian and constructivist principles are operating in visual processing, the debate centres on whether it is appropriate to ascribe these types of processing to discrete pathways. Whatever the outcome of the debate, Norman does present a compelling argument that visual processing does *not* have to be either for action *or* for recognition; it can be both.

## 6.5 Combining bottom-up and top-down processing

As we have shown, approaches to perception can be differentiated according to whether they are primarily concerned with perception for action or recognition, or with bottom-up or top-down processing. It may have occurred to you when reading about these approaches that it is likely that perception must in fact contain elements of both types of processing. A key question, then, is whether there is any evidence that this is in fact the case.

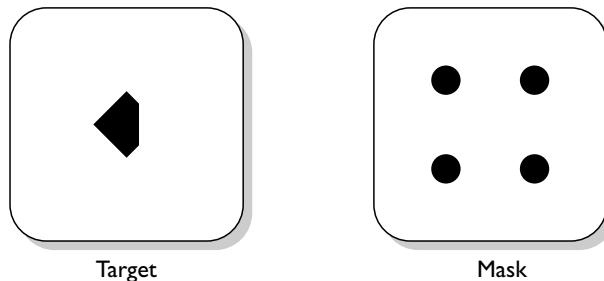
You were introduced to the idea of visual masking in the last chapter, particularly the concept of backward masking, in which the presentation of a second image disrupted the perception of an initial image. In Figure 3.26 you can see sets of stimuli that have been used to demonstrate two different types of visual masking. In each case, the mask is presented after a very brief presentation of the target. The task facing the participant is to report which corner of the diamond target is missing.



**Figure 3.26** Stimuli used to demonstrate backward masking

Standard explanations of why masking occurs with the stimuli in Figure 3.26 require that the mask contains contours that either overlap (Figure 3.26(a)) or exactly coincide with (Figure 3.26(b)) those of the target (Enns and Di Lollo, 2000). But, if masking is a product of the close similarity between the contours of target and mask,

it is hard to account for the fact that a masking effect is also found for the images in Figure 3.27 (Di Lollo *et al.*, 1993).



**Figure 3.27** An example of a four-dot mask

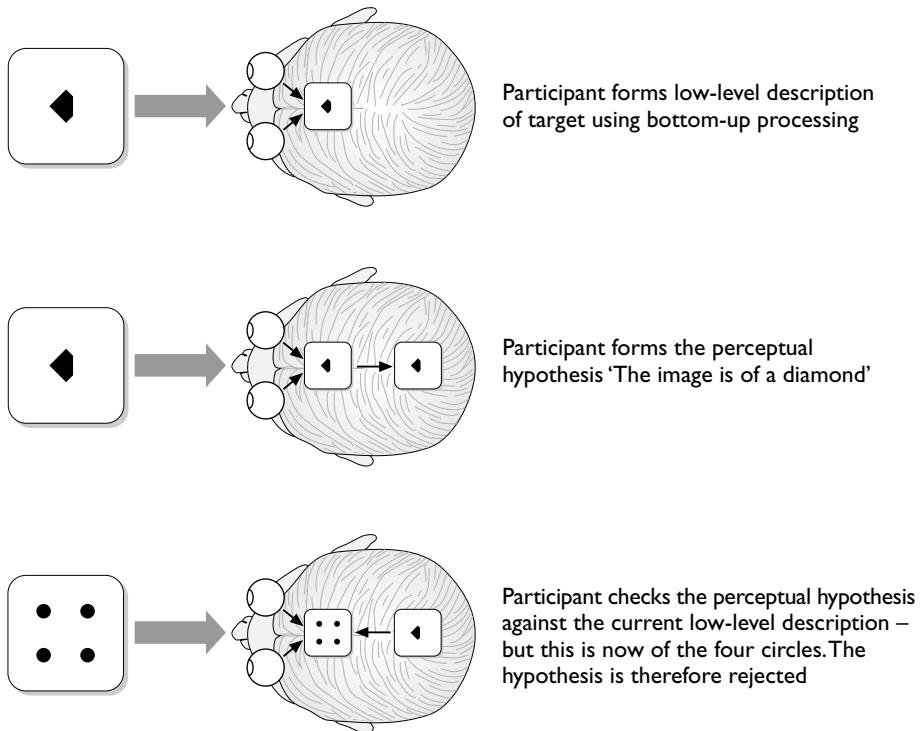
Enns and Di Lollo (1997) reported that the four-dot pattern shown in Figure 3.27 appeared to mask the target if target and mask were presented together and the target displayed very briefly, or if the mask was displayed very soon after a brief presentation of the target. Enns and Di Lollo (2000) explained the masking observed using the four-dot pattern by reference to **re-entrant processing**. We know from neuroscience research that communication between two different regions of the brain is never unidirectional. If one region is sending a signal to another, then the second region also sends a signal back through what are referred to as **re-entrant pathways** (Felleman and Van Essen, 1991).

Hupe *et al.* (1998) suggested that re-entrant pathways could be used to allow the brain to check a perceptual hypothesis against the information in an incoming signal. In other words:

- Bottom-up processing produces a low-level description.
- This is used to generate a perceptual hypothesis at a higher level.
- Using re-entrant pathways, the accuracy of the perceptual hypothesis is assessed by comparing it with the (perhaps now changed) low-level description.

Di Lollo *et al.* (2000) used this idea as the basis for an explanation of visual masking. The idea is that each part of the displayed image(s) is perceived in terms of a combination of high-level descriptions similar to a perceptual hypothesis and low-level codes produced by bottom-up processes. If the target is only presented very briefly, then masking can occur because by the time the high-level perceptual hypothesis is compared with the low-level bottom-up description, the target will have been replaced by the mask. Thus, the perceptual hypothesis will be rejected because it is based on a pattern (the target) that is different from the pattern currently being subjected to bottom-up processing (the mask) – see Figure 3.28.

The re-entrant processing explanation of visual masking is based upon the presumed interaction of bottom-up processes with top-down processes. This is consistent with the idea that perception is neither entirely bottom-up nor entirely top-down, but is actually reliant on both forms of processing.



**Figure 3.28** The re-entrant processing explanation of backward masking

## Summary of Section 6

---

- There appear to be at least two partially distinct, but interconnected streams of information flowing back from the retina to the primary visual cortex.
- From here, a ventral stream leads to the inferotemporal cortex and a dorsal stream to the parietal cortex.
- There is evidence that the ventral stream may be involved in perception for recognition and the dorsal stream in perception for action.
- Thus the dorsal stream would be better at dealing with the type of perception dealt with by Gibson and the ventral stream with the type of perception dealt with by Marr and the constructivist approach.
- Enns and Di Lollo's re-entrant processing explanation of backward masking was based on a combination of bottom-up and top-down perception.

## 7 Conclusion

We started this chapter by promising to show you just how complex even the perception of simple objects can be. We hope you now have some idea of these complexities and of the problems that face any potential theory of visual perception. You have also seen how rich the field of perception is. There are many influential

theories that have had a profound impact on both our understanding of perception and the way we approach cognitive psychology more generally. For example, Gibson showed us the importance of considering how we interact with the real world and Marr demonstrated the advantages of the modular approach to information processing. So, next time you are hunting in vain for your keys, do not be too hard on yourself. Remember all the computations, descriptions and hypotheses that your brain is having to process in order to perceive the environment around you.

## Further reading

- Bruce, V., Green, P.A. and Georgeson, M.A. (2003) *Visual Perception: Physiology, Psychology and Ecology*, Hove, Psychology Press.
- Gibson, J.J. (1986) *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates, Inc.
- Gregory, R.L. (1997) *Eye and Brain: The Psychology of Seeing*, Oxford, Oxford University Press.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman & Company.

## References

- Aksentijevic, A., Elliott, M.A. and Barber, P.J. (2001) ‘Dynamics of perceptual grouping: similarities in the organization of visual and auditory groups’, *Visual Cognition*, vol.8, pp.349–58.
- Atherton, M. (2002) ‘The origins of the sensation/perception distinction’, in Heyer, D. and Mausfeld, R. (eds) *Perception and the Physical World: Psychological and Philosophical Issues in Perception*, Chichester, John Wiley & Sons Ltd.
- Baizer, J.S., Ungerleider, L.G. and Desimone, R. (1991) ‘Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques’, *Journal of Neuroscience*, vol.11, pp.168–90.
- Binsted, G. and Carlton, L.G. (2002) ‘When is movement controlled by the dorsal stream?’, *Behavioral and Brain Sciences*, vol.25, pp.97–8.
- Bridgeman, B. (1992) ‘Conscious vs unconscious processing: the case of vision’, *Theory & Psychology*, vol.2, pp.73–88.
- Bridgeman, B., Peery, S. and Anand, S. (1997) ‘Interaction of cognitive and sensorimotor maps of visual space’, *Perception and Psychophysics*, vol.59, pp.456–69.
- Bullier, J. and Nowak, L.G. (1995) ‘Parallel versus serial processing: new vistas on the distributed organization of the visual system’, *Current Opinion in Neurobiology*, vol.5, pp.497–503.
- Courtney, S.M., Ungerleider, L.G., Keil, K. and Haxby, J.V. (1996) ‘Object and spatial visual working memory activate separate neural systems in human cortex’, *Cerebral Cortex*, vol.6, pp.39–49.

- Creem, S.H. and Proffitt, D.R. (1998) 'Two memories for geographical slant: separation and interdependence of action and awareness', *Psychonomic Bulletin and Review*, vol.5, pp.22–36.
- Di Lollo, V., Bischof, W.F. and Dixon, P. (1993) 'Stimulus-onset asynchrony is not necessary for motion perception or metacontrast masking', *Psychological Science*, vol.4, pp.260–3.
- Di Lollo, V., Enns, J.T. and Rensink, R.A. (2000) 'Competition for consciousness among visual events: the psychophysics of re-entrant pathways', *Journal of Experimental Psychology: General*, vol.129, pp.481–507.
- Enns, J.T. and Di Lollo, V. (1997) 'Object substitution: a new form of masking in unattended visual locations', *Psychological Science*, vol.8, pp.135–9.
- Enns, J.T. and Di Lollo, V. (2000) 'What's new in visual masking?', *Trends in Cognitive Science*, vol.4, pp.345–52.
- Enns, J.T. and Rensink, R.A. (1990) 'Sensitivity to three-dimensional orientation from line drawings', *Psychological Review*, vol.98, pp.335–51.
- Felleman, D.J. and Van Essen, D.C. (1991) 'Distributed hierarchical processing in primate visual cortex', *Cerebral Cortex*, vol.1, pp.1–47.
- Fitts, P.M. (1964) 'Perceptual-motor skills learning', in Melton, A.W. (ed.) *Categories of Human Learning*, New York, Academic Press.
- Gibson, J.J. (1947) 'Motion picture testing and research', *AAF Aviation Psychology Research Report No. 7*, Washington, DC, Government Printing Office.
- Gibson, J.J. (1950) *The Perception of the Visual World*, Boston, MA, Houghton Mifflin.
- Gibson, J.J. (1966) *The Senses Considered as Perceptual Systems*, Boston, MA, Houghton Mifflin.
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Goodale, M.A. and Milner, A.D. (1992) 'Separate visual pathways for perception and action', *Trends in Neurosciences*, vol.15, no.1, pp.20–5.
- Gregory, R.L. (1980) 'Perceptions as hypotheses', *Philosophical Transactions of the Royal Society of London, Series B*, vol.290, pp.181–97.
- Ho, C.E. (1998) 'Letter recognition reveals pathways of second-order and third-order motion', *Proceedings of the National Academy of Sciences of the United States of America*, vol.95, no.1, pp.400–4.
- Hupe, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P. and Bullier, J. (1998) 'Cortical feedback improves discrimination between figure and ground by V1, V2 and V3 neurons', *Nature*, vol.394, pp.784–7.
- Kanizsa, G. (1976) 'Subjective contours', *Scientific American*, vol.234, no.4, pp.48–52.
- Koffka, K. (1935) *Principles of Gestalt Psychology*, New York, Harcourt Brace.
- Kohler, W. (1947) *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*, New York, Liveright Publishing Corporation.
- Logothesis, N.K. (1994) 'Physiological studies of motion inputs', in Smith, A.T. (ed.) *Visual Detection of Motion*, London, Academic Press.

- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman & Company.
- Marr, D. and Hildreth, E. (1980) 'Theory of edge detection', *Proceedings of the Royal Society of London, Series B*, vol.207, pp.187–217.
- Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford, Oxford University Press.
- Milner, A.D. and Goodale, M.D. (1998) 'The visual brain in action', *Psyche*, vol.4, pp.1–14.
- Neisser, U. (1994) 'Multiple systems: a new approach to cognitive theory', *European Journal of Cognitive Psychology*, vol.6, no.3, pp.225–41.
- Norman, J. (2001) 'Ecological psychology and the two visual systems: not to worry', *Ecological Psychology*, vol.13, no.2, pp.135–45.
- Norman, J. (2002) 'Two visual systems and two theories of perception: an attempt to reconcile the constructivist and ecological approaches', *Behavioral and Brain Sciences*, vol.25, no.1, pp.73–96.
- Penrose, L.S. and Penrose, R. (1958) 'Impossible objects: a special type of illusion', *British Journal of Psychology*, vol.49, p.31.
- Rao, S.C., Rainer, G. and Miller, E.K. (1997) 'Integration of what and where in the primate prefrontal cortex', *Science*, vol.276, pp.821–4.
- Rock, I. (1977) 'In defense of unconscious inference', in Epstein, W. (ed.) *Stability and Constancy in Visual Perception: Mechanisms and Processes*, New York, Wiley.
- Rock, I. (1983) *The Logic of Perception*, Cambridge, MA, MIT Press.
- Rock, I. (1997) *Indirect Perception*, Cambridge, MA, MIT Press.
- Schneider, G.E. (1967) 'Contrasting visuomotor functions of tectum and cortex in the golden hamster', *Psychologische Forschung*, vol.31, pp.52–62.
- Schneider, G.E. (1969) 'Two visual systems', *Science*, vol.163, no.3870, pp.895–902.
- Sedgwick, H.A. (1973) *The Visible Horizon*, Unpublished PhD thesis, Cornell University Library.
- Shapley, R. (1995) 'Parallel neural pathways and visual function', in Gazzaniga, M.S. (ed.) *The Cognitive Neurosciences*, Cambridge, MA, MIT Press.
- Street, R.F. (1931) *A Gestalt Completion Test*, New York, Bureau of Publications, Teachers College, Columbia University.
- Ungerleider, L.G. and Mishkin, M. (1982) 'Two cortical visual systems', in Ingle, D.J., Goodale, M.A. and Mansfield, R.J.W. (eds) *Analysis of Visual Behaviour*, Cambridge, MA, MIT Press.
- Wade, N.J. and Bruce, V. (2001) 'Surveying the scene: 100 years of British vision', *The British Journal of Psychology*, vol.92, no.1, pp.79–113.
- Wertheimer, M. (1923) 'Untersuchungen zur Lehre von der Gestalt, II', *Psychologische Forschung*, no.4, pp.301–50. Translated as 'Laws of organization in perceptual forms', in Ellis, W.D. (ed.) (1955) *A Source Book of Gestalt Psychology*, London, Routledge and Kegan Paul.

Young, M.J., Landey, M.S. and Maloney, L.T. (1993) 'A perturbation analysis of depth perception from combinations of texture and motion cues', *Vision Research*, vol.33, pp.2685–96.

# Recognition

## Chapter 4

Graham Pike and Nicola Brace

---

### 1 Introduction

In the last chapter on perception, we explored some of the cognitive processes involved in forming a mental description of the environment based on input from the senses. As well as being able to determine the position and shape of the objects around us, it is also possible to recognize *what* we are seeing. Unless we fully accept Gibson's concept of affordance (and it's safe to say that we don't), there must be another step: another set of processes that transform the basic descriptions of objects generated by analysing the retinal image into objects that are familiar to us and which we can recognize.

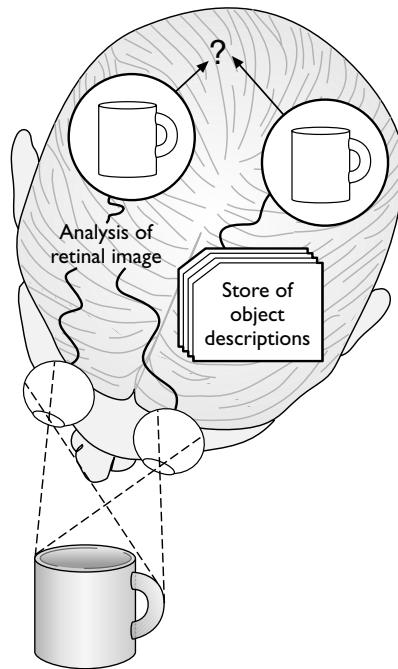
The same is, of course, true of our other senses; for example, when we listen we may hear music, car engines and voices. Again, there must be cognitive processes that somehow transform the auditory input of sound waves into what we recognize as an environment of voices, music and cars.

Let's stop for a moment and consider the basic steps that might be involved in the process of visually recognizing an object:

- First, there must be processes that are able to construct an internal representation (referred to as a 'description') of the object, based on the information in the retinal image.
- Second, there must be processes that are able to store this description so that we can recognize the object if we see it again.
- Third, there must be processes that somehow compare the description of the object that we can currently see to the descriptions of objects that we have stored.
- Lastly, it is very likely that we have seen objects from many different angles, yet are able to recognize them regardless of the current angle of view. As we shall see, the nature of the mechanism that allows us to do this is an important and much debated point.

A basic diagram displaying the recognition process is provided in Figure 4.1 (overleaf).

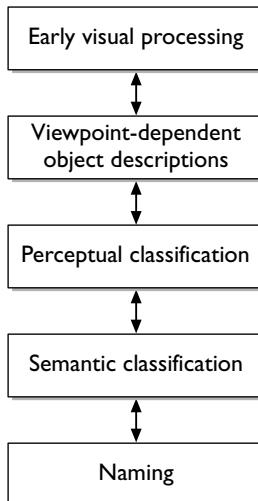
In one sense, the process of recognition is the process of generating and comparing descriptions of objects that are currently in view with descriptions of objects that we have seen previously. It is worth noting that this is a very simplistic way of viewing and describing recognition, and in Section 2 we shall look at some of the problems with this simplistic approach.



**Figure 4.1** The basics of the recognition process

### 1.1 Recognition in the wider context of cognition

In Figure 4.2, we can see how Humphreys and Bruce (1989) summarized the way in which object recognition fits into a wider context of cognition that includes perception (perceptual classification), categorization (semantic classification) and naming. As you can see from Figure 4.2, the first stage in the process is the early visual processing of the retinal image. One example of this form of processing is that which produces Marr's full primal sketch (Marr, 1982). In the second stage a



**Figure 4.2** Model of object recognition suggested by Humphreys and Bruce (1989)

description of the object is generated, but this description is dependent on the viewpoint of the observer. This stage is therefore similar to what Marr (1982) referred to as the 2½D sketch.

Humphreys and Bruce refer to the next stage as ‘perceptual classification’ and it is really this stage that we have been discussing so far in this chapter. Perceptual classification involves a comparison of the information regarding the object in view with descriptions of objects that have been stored previously. It is at this stage that the object is ‘recognized’.

Once the object has been recognized, or perceptually classified, it can then be ‘semantically classified’. This process, also referred to as ‘categorization’, is examined in the next chapter. Once this stage has been achieved, the object can then be named, aspects of which will be examined in the later chapters on language.

## Summary of Section 1

---

- As well as being able to determine the location and shape of an object, or the location and pitch of a sound, we also have to be able to recognize what they are.
- A basic model of recognition requires that a description from sensory input is generated and compared with descriptions stored in memory.
- Recognition must come after the initial processes of perception and before the stages in which an object can be first semantically classified and then named.

## 2 Different types of recognition

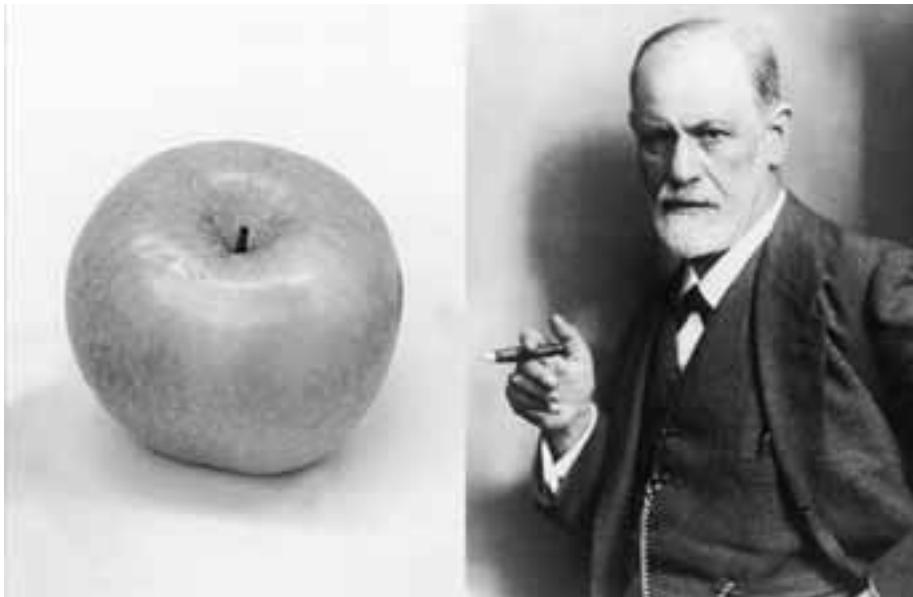
As we have stated above, the view that recognition involves comparing an object description generated from the retinal image to descriptions stored in long-term memory is very simplistic. In fact there are quite different *types* of recognition, depending on what it is we are trying to recognize and how we go about trying to recognize it. Throughout this chapter we shall be exploring these different types of recognition and examining some of the issues that suggest the process of recognition is far more complex than the simplistic model presented in Figure 4.1 suggests.

### 2.1 Object and face recognition

The end point of Humphreys and Bruce’s (1989) model of recognition (Figure 4.2) is the naming stage. Naming, of course, is not a necessary component of being able to recognize an object: even if an animal has no capacity for language, it can still recognize objects. But the names we give things do provide a clue to the fact that there are different types of recognition.

**ACTIVITY 4.1**

Figure 4.3 shows two images. See if you can name them.



**Figure 4.3**

You probably provided the names ‘apple’ and ‘Sigmund Freud’. These are evidently two different types of name, but can you describe why these two types of name are so different? Hint: think about how many different apples and Sigmund Freuds there are.

In completing Activity 4.1, you may have realized that the name you provided for the left-hand image was the category to which the object belonged, whilst the name for the right-hand image corresponded to an individual rather than a category (i.e. you did not name the image ‘a face’).

Naming reveals that it is possible to recognize objects in different ways. When we see objects such as fruit and furniture we tend to concentrate on which category they belong to, and when we provide names for them, these are usually the name for that category. Thus, we are making **between-category** distinctions such as ‘that object is an orange and that one is a table’. However, when we see a face, we often do more than recognize that the object belongs to the category of objects known as ‘faces’, we also work out *whose* face it is. In other words, we make a **within-category** distinction.

The difference between within- and between-category recognition is one reason why **face recognition** is generally researched as a separate topic from **object recognition**. In addition, there are some issues that are unique to face recognition such as:

- The internal features of a face can move, which changes the appearance of the face.

- This movement can serve to express emotional and social cues.
- Faces can change quite dramatically over time, due to ageing or haircuts for instance.

## ACTIVITY 4.2

---

Can you identify the person depicted in the three images shown in Figure 4.4?



**Figure 4.4**

### COMMENT

The images are of Paul McCartney and you were (probably) able to recognize him from all three images, even though there are some quite obvious differences in appearance. In fact, you were probably able to recognize the E-FIT image of him (right-hand image), even though this is constructed by combining together features from several other faces. So, we can recognize a face that is familiar to us even when quite large changes have been introduced.

---

As well as distinguishing between face and more general object recognition, it is possible to identify a number of different types of face recognition. One such distinction is between recognizing familiar and unfamiliar faces. Pike *et al.* (2000) reported that people were often able to identify E-FIT images even when other participants had rated them as a poor likeness. However, like the E-FIT in Activity 4.2, these were images of famous people, whose faces would have been familiar to the participants. Considerable evidence suggests we are not so accurate at recognizing even real faces that are not so familiar to us. For example, many witnesses express uncertainty when asked to identify the perpetrator of a crime from a line-up (Pike *et al.*, 2001). Even when the anxiety of the witness is reduced by using a video identification parade, identification accuracy is far from perfect (Kemp *et al.*, 2001).

A second distinction that applies to types of face recognition is that between recognizing whose face you are looking at and recognizing what emotion it may be portraying. You can imagine that the importance of faces in conveying emotional state and in facilitating social interactions has led us to develop some very sophisticated cognitive processes for interpreting facial expressions. In fact, we are

able to judge the emotion being displayed on a face with great accuracy (the cognition involved in perceiving emotion is considered in Chapter 13) and are very sensitive to eye movements in those around us. It is tempting to think that we may have evolved a specific set of cognitive processes for recognizing faces and the emotions they express because of the social importance of this information. However, there is evidence (Young *et al.*, 1993) that although we do have specific processes for recognizing emotions, these processes are not involved in recognizing identity. We shall return to the difference between emotion and identity recognition later in this chapter, but logically you can see that you need to be able to tell whether someone is angry or happy regardless of whether you can recognize them or not. Likewise, you need to be able to recognize who someone is regardless of whether they appear happy or angry.

The question of whether faces are recognized by the same cognitive processes that are used to recognize other objects has been at the centre of a great deal of research. Although a definitive answer as to just how different face recognition is from general object recognition has yet to be provided, the two have tended to be treated as different areas of research. Because of this, we have divided this chapter into two main areas of discussion. The first (the rest of Section 2 and Section 3) will look at theories of how we recognize objects, and the second (Sections 4 to 7) will look at models of face recognition and examine in more depth the question of whether faces are recognized by special processes.

## 2.2 Active processing – recognizing objects by touch

One limitation of the basic recognition procedure we suggested in Section 1 is that it treats recognition as a passive process. Gibson (1986) stressed that perception is an active process and that we are beings who interact with and investigate the environment. In examining how Gibson's idea of active perception might apply to recognition, we will temporarily switch modalities from vision to touch. One reason for concentrating on touch is that purely passive object recognition through touch would be almost impossible. Although there may be some objects that you can recognize if they were simply placed on your hand, most objects would require exploration. We have evolved sophisticated processes for exploring the environment and objects using touch in very exact and careful ways.

First, we have tremendous control over our hands, so that we can both move our fingers precisely and also apply varying degrees of pressure to objects in a very measured way. This is done by employing a **feedback** system, whereby information from touch receptors allows the brain to control the location and amount of pressure applied by the fingers. As well as being able to regulate touch precisely, we can also pinpoint the location of our limbs with great accuracy via receptors inside our muscles and joints. This information about limb location is known as **kinesthesia**, and it can be combined with information from the touch receptors to guide our hands and fingers. Of particular importance are the relative positions of your fingers as they touch the object, their orientation to your hand, and the position of your hand in relation to your arm and of your arm in relation to your body. The processes that allow us to keep track of the relative locations of all our limbs are known as **proprioception**.

So, at every moment that we are touching an object, we know the exact position of our fingers (kinesthesia) and what the object feels like at that point (touch receptor information). The information gained from this combination is referred to as **haptic information** and it can be used to generate a description of an object.

Lederman and Klatzky (1987) found that there was considerable consistency in the way in which people used their hands in order to gather haptic information. They described how participants tended to use a series of **exploratory procedures** when investigating an object with their hands. Lederman and Klatzky (1990) went on to study these exploratory procedures in more depth and described how each particular procedure could be used to derive a certain type of information that was useful for recognizing an object. For example, if shape was important in recognizing the object people tended to move their fingers around the object's contours, and if texture was important they would move their fingers across the surface of the object.

### ACTIVITY 4.3

---

Ask someone to place a variety of objects within easy reach of you (you can do this yourself if you wish). Ask them to choose objects that differ in shape, texture and weight. Close your eyes and pick up each object in turn and try to work out what it is. As you do this, try to make a mental note of the different movements that your hands make and what each movement tells you about the object.

Table 4.1 gives a list of some of the hand movements reported by Lederman and Klatzky (1987), along with the information that these exploratory procedures tend to reveal. Did you find yourself using these movements?

**Table 4.1** The information revealed by exploratory hand movements

Movement	Information
Enclose object in hand(s)	Overall shape
Following contours with fingers	More exact shape
Lateral motion with fingers	Texture
Press with fingers	Hardness
Static contact with fingers	Temperature
Unsupported holding	Weight

Source: based on Lederman and Klatzky, 1987, Table 1, p.345

---

Although haptic perception can be used to recognize objects, visual recognition has the obvious advantage that it can be used for distant objects that are out of reach and tends to be far quicker and more accurate in processing information about shape, particularly complex 3D shape (Lederman *et al.*, 1993). But, visual perception is not so useful when it comes to judging the weight and texture of an object.

So, haptic perception is a very useful source of information and can be used to recognize certain objects. The study of haptic perception also serves to demonstrate that recognition is not necessarily passive and that much can be gained from

considering it as an active process. Nor is active perception limited to touch. You saw in the last chapter how your interpretation of the impossible triangle (Figure 3.24) kept changing as you visually explored the object, corner by corner.

## 2.2 Recognizing two-dimensional objects

Another way of distinguishing between types of recognition is according to whether the object in question is three-dimensional (3D), such as the book in front of you, or two-dimensional (2D), such as the words in front of you. The difference between 2D and 3D object recognition takes on added significance when you consider that the description generated from the retinal image will in essence be 2D, whilst most objects tend to be 3D. In fact, much of the early research conducted on recognition processes was focused on how simple, two-dimensional ‘patterns’ are recognized. Although it can be argued that this work tells us little about how complex, three-dimensional objects are recognized, it does serve to highlight some of the problems that are inherent in any approach to object recognition.

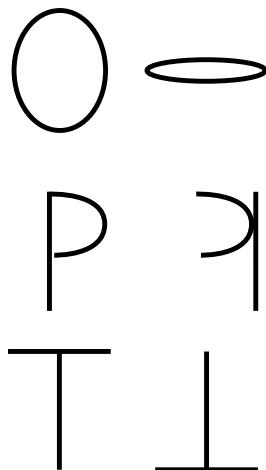
By far the simplest model of visual pattern recognition postulates **template matching**. This is the idea that we have a large number of templates stored in long-term memory against which we compare the patterns we come across. For example, a template would exist for every number from 0 to 9 and for every letter from A to Z. The problem with this theory is that it cannot cope with the enormous variation in the actual patterns that are used to represent even simple things such as alphanumeric characters. For example, in Figure 4.5 the top row contains examples of the letter ‘R’ and the bottom row contains examples of letters, each of which shares many similar properties with the specific example of an ‘R’ immediately above it. Although we do not have any great difficulty in reading these letters, it is hard to see how a simple template could be created that would accept every example in the top row as a letter ‘R’ and reject every example in the bottom row.



**Figure 4.5** Different alphanumeric characters that share similar properties

If the problem with template matching is that the template cannot deal with variation in the stimulus it has to recognize, we have to look at some way of representing objects that is not so reliant on the exact visible pattern. One way of doing this is to try to extract the key characteristics or features of an object. In the case of alphanumeric characters, these features could be the number of curved and straight lines and the relationship between them. An ‘O’ might therefore be represented as a single continuous curve, a ‘P’ as one vertical line and one discontinuous curve, and a ‘T’ as one horizontal and one vertical line that form two right angles.

One of the most influential **feature recognition theories** is the Pandemonium system, so called because processing units known as ‘demons’ were used to detect each feature. This system was designed as the basis for a computer program to decode Morse code signals (Selfridge, 1959) and was later adapted by Neisser (1967) to recognize alphanumeric characters. Although Pandemonium systems have been useful in recognizing simple, highly constrained patterns, they do not provide a particularly useful model of human object recognition. A central flaw in feature recognition theories is that describing an object in terms of a list of key features does not capture the structural relations *between* features. If you look back at the feature-based descriptions provided for an ‘O’, ‘P’ and ‘T’ above, you will see that these three descriptions could also apply to the figures presented to the right of each letter in Figure 4.6, meaning that these shapes would be misidentified as letters.



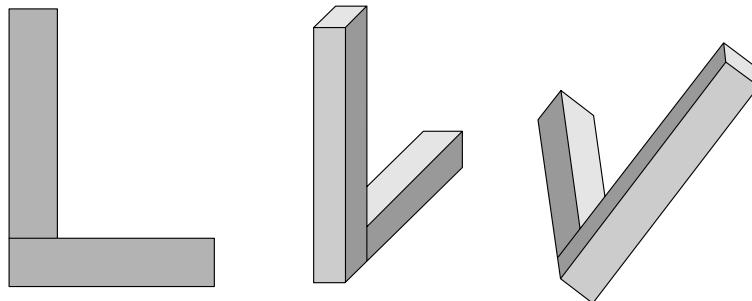
**Figure 4.6** Examples of different patterns described by the same key features

An approach that has had more success in explaining how both simple patterns and more complex objects might be recognized is that based on **structural descriptions**. Structural descriptions are made up of a series of propositions, based both on a description of the elements that comprise the object and the structural relations between them. Thus, the structural description of a letter ‘L’ might contain the following propositions:

- There are two lines.
- There is one horizontal line.
- There is one vertical line.
- The horizontal line supports the vertical line.
- The horizontal line extends to the right of the vertical line.
- The horizontal and vertical lines are joined at a right angle.

Although the propositions stated above are expressed in language, they can be equally well expressed in other forms of symbolic representation, such as that used in a computer program.

One key advantage that structural descriptions have is that it is possible to see how they could be applied to three-dimensional objects. Consider the three representations of the character 'L' in Figure 4.7. Both template matching and feature recognition theories would recognize the representation to the left as being an 'L', but would immediately reject the other two. However, the two forms of the letter 'L' on the right of Figure 4.7 do share a similar structural description once we consider their three-dimensional properties.

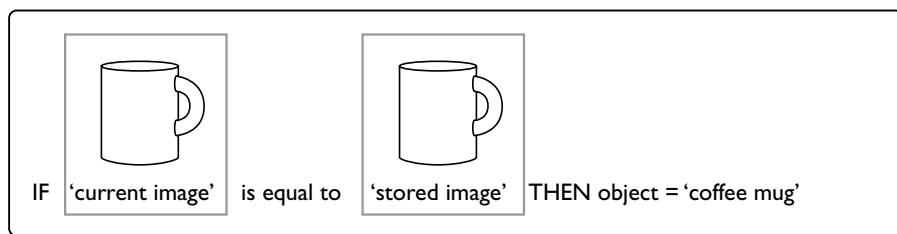


**Figure 4.7** Three representations of a 3D 'L' shape

But, in order to obtain a description that includes elements of three-dimensional structure, we must be able to turn the 2D retinal image, that is dependent on the particular view that the observer has of the object, into a 3D description that is centred not on the viewer but on the object itself. This, as you might expect, requires an even more sophisticated means of describing objects, and is the focus of the second half of Marr's theory of vision – which we shall look at in Section 3.

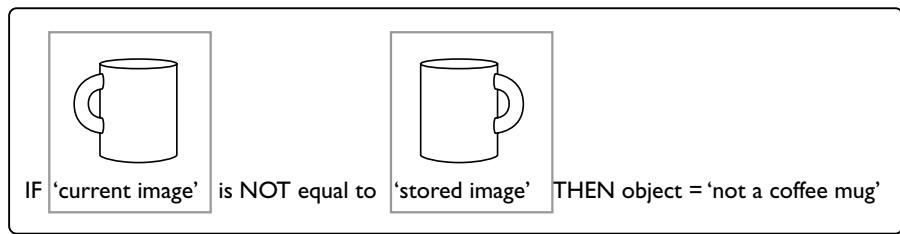
### 2.3 Object-centred vs viewer-centred descriptions

One of the most fundamental problems in recognizing an object is that it is possible to view an object from many angles. As we have seen, any theory that treats an object as a simple pattern is likely to fail when applied to a 3D object (as with the 'L' in Figure 4.7). Consider writing a very simple computer program based on recognizing an object by matching patterns. As an example, Figure 4.8 contains a conceptual diagram of how a computer might be programmed to recognize a coffee mug.



**Figure 4.8** A simple program for recognizing an object

But coffee cups are actually 3D objects and can be viewed from many angles. Let's see how our simple computer program would cope if we turned our coffee cup so it was facing the other way. As you can see from Figure 4.9, the program has decided that, as the patterns do not match, the object is NOT a coffee cup.



**Figure 4.9** A simple program failing to recognize an object from a different viewpoint

The failure of the simple program to deal with a small change in viewpoint is obviously an unacceptable flaw in any system that wishes to interact with its environment. Instead of being reliant on seeing objects from just a single viewpoint, the process of object recognition must somehow be based on descriptions of objects that allow recognition to take place *independent* of viewpoint. In fact, these processes must be tolerant of any naturally occurring change, not just changes in viewpoint. This is a very important point and one that is central to the study of object recognition.

Marr (1982) conceptualized the problem of viewpoint as that of turning the **viewer-centred description** of the object that was formed in the 2½D sketch (see Chapter 3, Section 4.3) into a **3D object-centred description** that would allow the object to be recognized despite changes in viewpoint. In the next section we shall look at how Marr suggested this might happen.

## Summary of Section 2

---

- There are different types of recognition, that depend on what is being recognized and how.
- Object recognition tends to be based on making between-category distinctions and face recognition on making within-category distinctions.
- Face recognition tends to be researched apart from more general object recognition because faces can convey social and emotional information and their appearance can change.
- Recognition is not entirely a 'passive' process and can involve an active exploration of the environment. This is particularly true of haptic recognition, in which objects are recognized by touch.
- One key problem facing any theory of visual recognition is that the retinal image is essentially 2D, but objects are 3D.
- Early theories that concentrated on recognizing 2D patterns, such as template matching and feature recognition theories, are therefore not particularly useful models of human recognition.
- Theories based on abstracting a structural description of an object are better able to cope with 3D objects.
- As a 3D object can be viewed from many angles, our recognition system must be able to turn an object description centred on the viewer into one centred on the object.

### 3 Recognizing three-dimensional objects

As we saw in the previous chapter, in the first part of Marr's theory of perception, early visual processing of the retinal image eventually leads to the generation of the 2½D sketch. But the surfaces and objects in the 2½D sketch are described in relation to the viewpoint of the observer and are therefore viewer-centred descriptions. As we saw in the previous section, viewer-centred descriptions are of little use in recognizing real objects that can be seen from any angle and any distance. The second half of Marr's theory was therefore concerned with how the information in the 2½D sketch might be used in order to construct a 3D object-centred description of each object.

If it were not possible to generate a 3D object-centred description, the only way of accurately recognizing objects would be to store a very large number of viewer-centred descriptions. Although there are theories that have taken this approach, for now we will concentrate on the idea that recognition is best subserved by a single representation of an object that can be used to recognize it from any angle.

Marr and Nishihara (1978) suggested that objects could be represented by generating a 3D object-centred description that would allow the object to be recognized from virtually any angle. They proposed that this description was based on a **canonical coordinate frame**. This basically means that each object would be represented within a framework that was about the same shape as the object. You could imagine the representation of a carrot as being a cylinder that tapered toward one end.

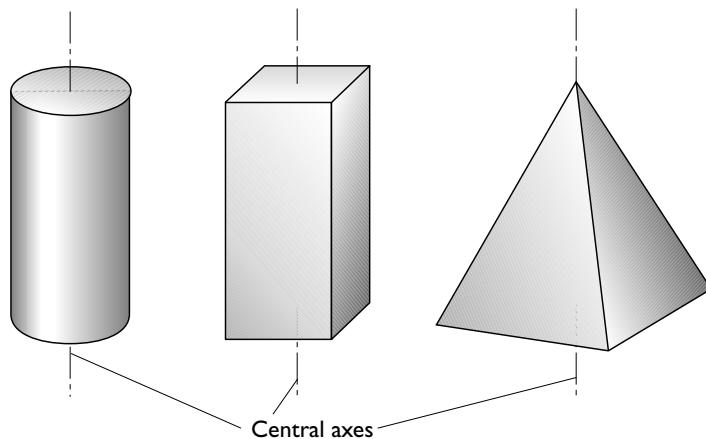
This procedure appears at first glance to be somewhat paradoxical, as it would be necessary to know the approximate shape of the object before you could begin to recognize it! However, remember that the formation of the 3D object-centred description occurs after considerable analysis of the retinal image has already taken place, so some information as to the shape/outline of the object will already exist.

#### 3.1 Marr and Nishihara's theory

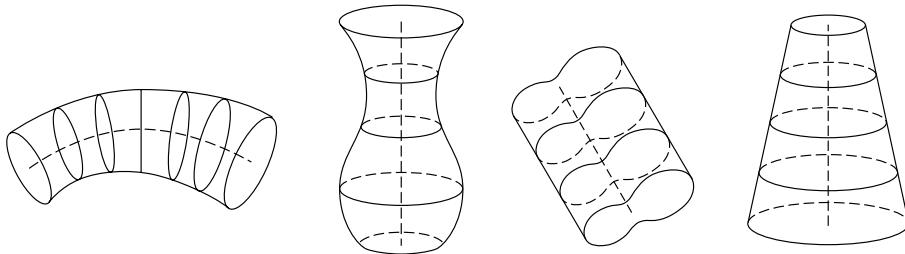
Marr and Nishihara saw the first step in establishing a canonical coordinate frame as defining a central axis for the object in question. This is relatively easy to do if the object in question either has a natural line of symmetry or has a length that is noticeably greater than its width and depth (see Figure 4.10).

In fact, the generation of the central axis is so important in Marr and Nishihara's theory that it is restricted to specific objects that can be easily described by one or more **generalized cones**. A generalized cone is any 3D shape that has a cross-section of a consistent shape throughout its length. The cross-section can vary in size, but not in shape. All of the objects shown in Figure 4.11 are examples of generalized cones. Although restricting the theory to generalized cones is undoubtedly one weakness of Marr and Nishihara's theory, the basic shape of many natural objects, particularly those that grow (such as animals and plants), can be described, albeit rather loosely, in this way.

To locate the central axis of an object, it is first necessary to make use of the information contained within the 2½ D sketch in order to work out what shape the object has. Marr (1977) suggested that it is possible to work out the shape of an



**Figure 4.10** Locating the central axis of an object



**Figure 4.11** Three generalized cones

Source: Marr, 1982, Figure 3.59, p.224

object based on the object's **occluding contours** (these are basically the object's silhouette). The points on the object's surface that correspond to the boundary of its silhouette are of particular importance in Marr's theory, and he referred to them as the **contour generator** – because they can be used to generate the contour of the object.

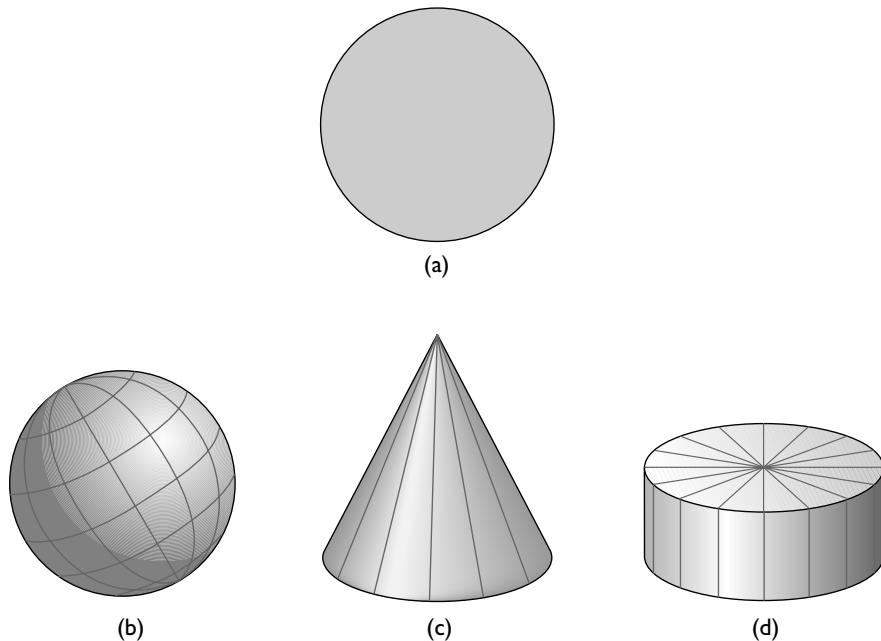
As Marr (1982) points out, we seem to have no problems in deriving 3D shapes from silhouettes such as those used in Picasso's *Rites of Spring* (see Figure 4.12).

However, as the silhouette of an object is two-dimensional, it is possible that it could be caused by more than one 3D object. Consider the circular silhouette (a) in Figure 4.13. This could be caused by any of the 3D objects below it (if they were sufficiently rotated), yet we tend to interpret the silhouette as being produced by the sphere (b).

Marr suggested that the problem of how we can derive 3D shape from 2D silhouettes is solved by the visual system making certain assumptions about what it is seeing. As Marr himself said, 'Somewhere buried in the perceptual machinery that can interpret silhouettes as three-dimensional shapes, there must lie some source of additional information that constrains us to see the silhouettes as we do' (Marr, 1982, p.219). Marr conceptualized this 'additional information' as coming in the form of three basic assumptions built into the computational processes:



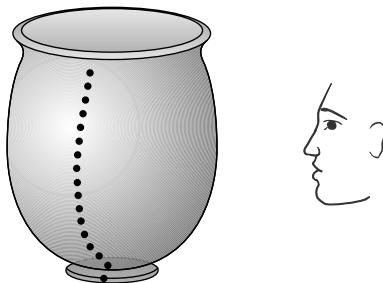
**Figure 4.12** *Rites of Spring* by Picasso



**Figure 4.13** A silhouette (a) and three objects that could cause it (b, c and d)

- Each point on the contour generator corresponds to a different point on the object.
- Any two points that are close together on the contour in an image are also close together on the contour generator of the object.
- All the points on the contour generator lie in a single plane (i.e. are planar).

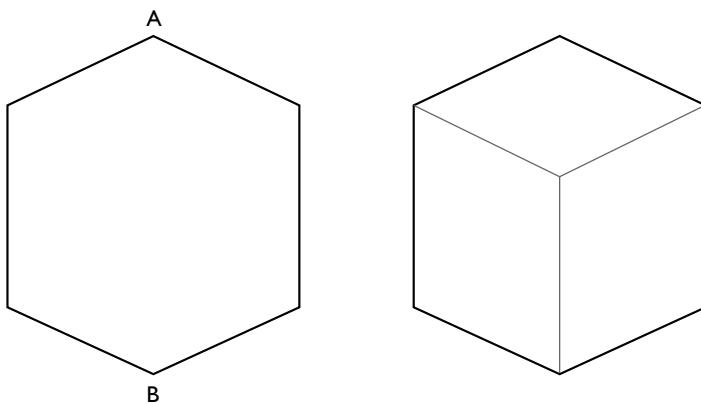
The first two points are relatively straightforward and the third assumption has been illustrated in Figure 4.14.



**Figure 4.14** The black dots indicate points that lie in the same plane with respect to the viewer

Source: Marr, 1982, Figure 3.57(d), p.220

The third assumption, that all of the points on the contour generator are planar, is a vital component in Marr's theory, but it can be problematic. As we have seen, it is possible for two quite different objects to share the same silhouette and for the points on the silhouette to vary in their distance from the observer. We tend to interpret the contour on the left in Figure 4.15 as being a hexagon. However, this contour will be produced by the cube to the right. The problem is that the assumption that all the points on the contour are planar is violated by this view of the cube, as point (A) is further away than point (B). As the points on the cube's occluding contour are not planar, we tend to interpret its silhouette incorrectly.



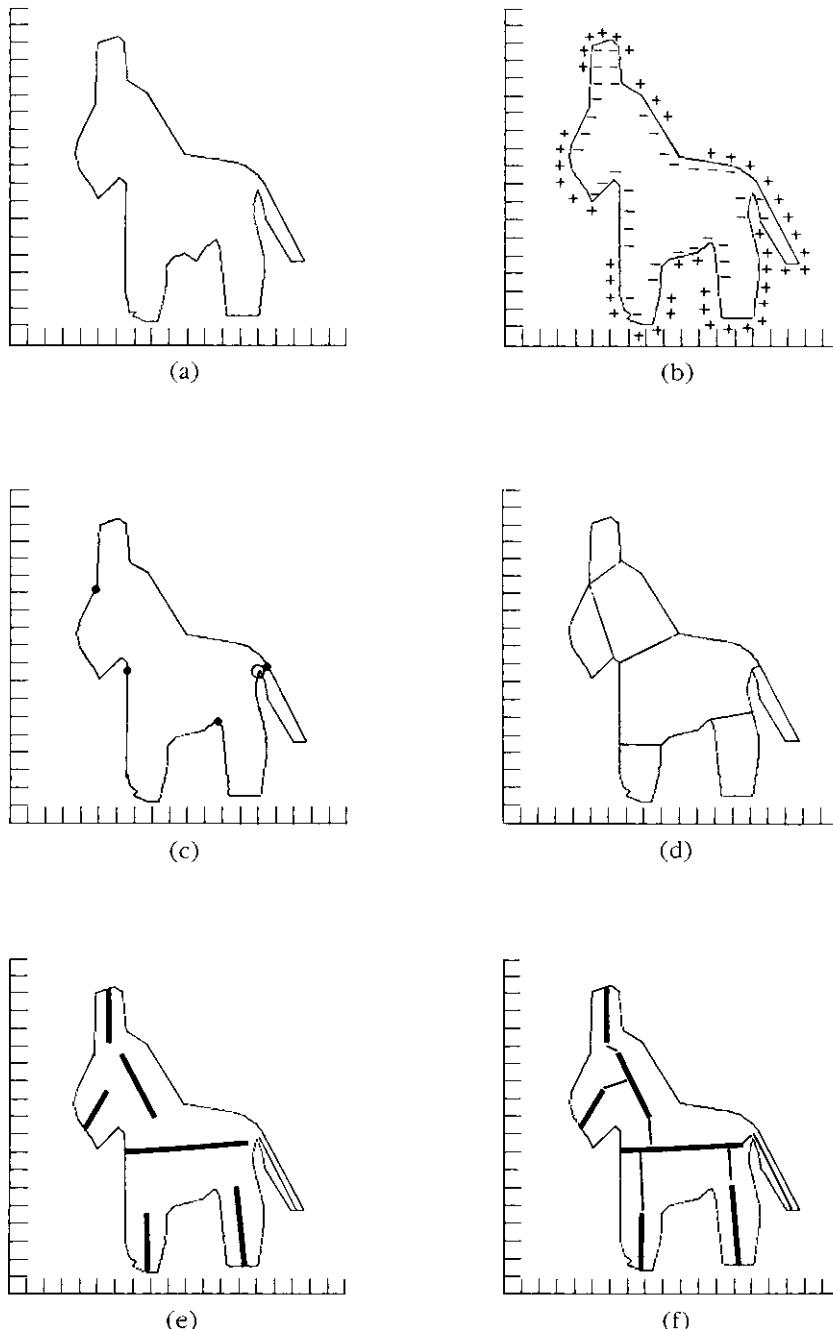
**Figure 4.15** The contour of a cube may not be planar

Source: based on Marr, 1982, Figure 3.58, p.221

Once the shape of the object has been derived using its contour generator, the next step is to locate the axis/axes necessary to represent the object. It is fairly straightforward to do this when the shape is simple as symmetry usually tells us where its axis is located, but what about more complex shapes? The answer is that we often need to represent the shape using several axes, so that the object is divided into components and one axis is used for each component (these are referred to as **component axes**).

In Figure 4.16, one method of locating axes suggested by Marr and Nishihara (1978) is illustrated. As you can see, the object in question is a toy donkey (a). The first step (b) involves working out areas of **concavity** (these correspond to parts of

the contour that include a bend inwards and are represented in the figure by a ‘–’) and **convexity** (parts of the contour that include a bend outwards and are represented by a ‘+’). The shape can then be divided into sections by finding areas of sharp concavity (c) and using these to divide the object into smaller parts (d). Once the shape has been divided in this way, it is possible to represent each section via a component axis (e).



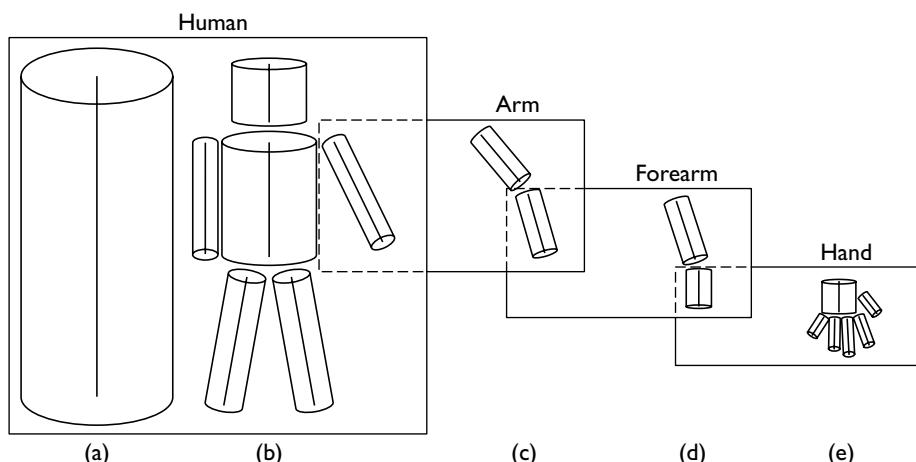
**Figure 4.16** Locating the component axes of an object

Source: Marr and Nishihara, 1978, Figure 6

These component axes can then be represented in relation to the horizontal axis of the body (f).

Figure 4.17 illustrates how it is possible to represent a quite complex object using several components or **primitives** as Marr called them. The description of the object must allow recognition at a global level, such as being able to tell that an object is a human body, and also incorporate more detailed information, such as the fact that a human hand has five fingers. It is therefore necessary for there to be a hierarchy of 3D models, in which each subsequent level contains a more detailed description of a specific part of the object. This means that fewer primitives will be used to represent each part of the object at the higher levels of the hierarchy.

For example, consider the description of the human body provided in Figure 4.17. At the highest level, the entire human body is described in relation to a single axis that runs through the centre of the body (a). This 3D model also contains the relative length and orientation of the axes that describe the head, torso, arms and legs (b). However, no details regarding smaller parts (such as the fingers) are provided.



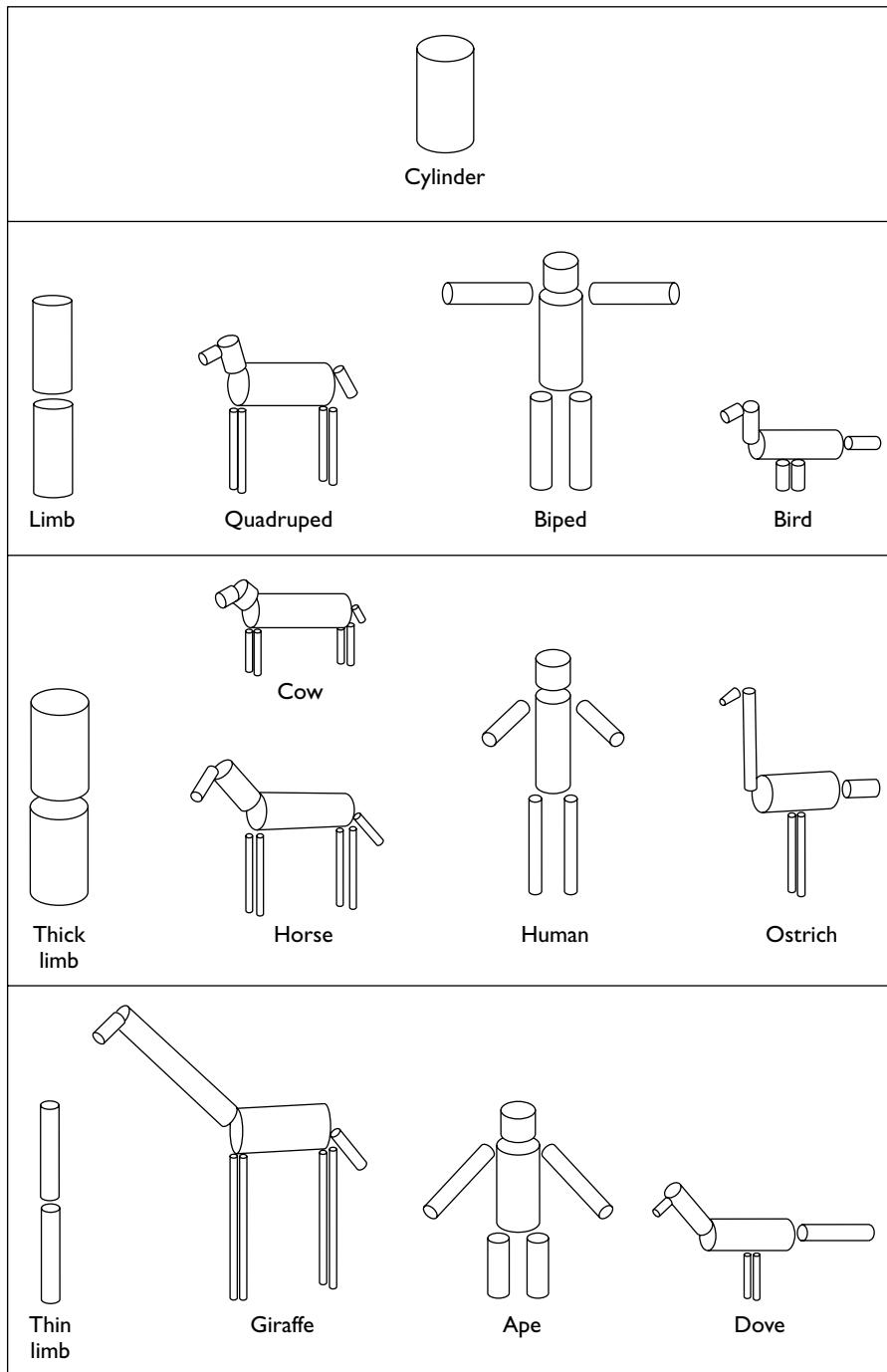
**Figure 4.17** Marr and Nishihara's hierarchical model of a human body

Source: Marr and Nishihara, 1978, Figure 3

The axis that corresponds to each limb (b) is then used as the major axis for a more detailed description of that limb (c). For example, the axis of the cylinder representing the right arm is then used as the major axis to represent the upper and lower part of that arm (c). The axis of the cylinder used to describe the lower part of the arm (c) is used as the major axis to describe the forearm and hand (d). Finally, the axis of the cylinder used to describe the hand (d) is used as the major axis in order to describe the five fingers (e). Thus we have a 3D model description that can be used to recognize an entire human body as well as any of its parts.

Having derived a 3D description of the object, Marr and Nishihara (1978) saw the next step in the process of recognition as comparing this to a **catalogue of 3D models**, formed from the 3D descriptions of all previously seen objects. The catalogue is organized hierarchically according to the amount of detail present in the model (see Figure 4.18). At the highest level the catalogue consists of descriptions devoid of any decomposition into components. The next level contains more detail,

corresponding to the number and basic layout of limbs as in Figure 4.17. At the next level even more detail is contained, such as that relating to the angles and lengths of component axes.



**Figure 4.18** 3D model catalogue

Source: Marr and Nishihara, 1978, Figure 8

The 3D model generated of a new object (the target) is related to the catalogue, starting at the highest level. The target is compared to the stored models and the example it best matches is used as the basis for the next level of detail. The process stops when a level is reached that corresponds to the level of detail present in the target. At this point, assuming the target contains sufficient detail, a match should have been found that allows the object to be recognized.

So, the generation of a 3D model description solves several problems inherent to object recognition. As the model is 3D, it allows recognition of the object from many angles and its hierarchical nature allows recognition of the entire object whilst maintaining more detailed information about the components.

### **3.2 Evaluating Marr and Nishihara's theory**

Although it can be difficult to study the cognition involved in object recognition, there is evidence for some of the suggestions made by Marr and Nishihara.

One of the key predictions of their theory arises from the fact that they see establishing a central axis as a vital stage in the recognition process. This means that it should be very difficult to recognize an object if it is also difficult to establish the location of its central axis. Some support for this notion comes from a study, conducted by Lawson and Humphreys (1996), in which participants had to recognize objects (line drawings in this case) that had been rotated. Rotation did not appear to have an effect on recognition unless the major axis of the object was tilted toward the observer. Presumably, the disruption to recognition was due to the major axis appearing foreshortened and therefore harder to locate.

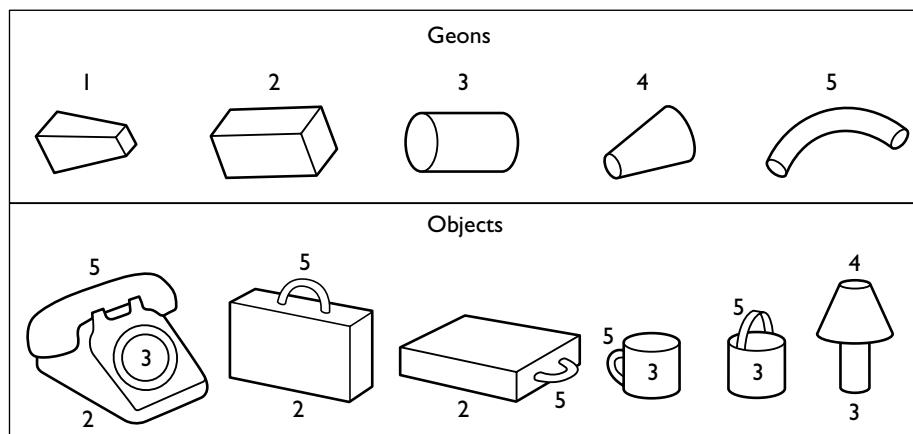
More powerful evidence in support of Marr and Nishihara's theory comes from neuropsychological case studies. Warrington and Taylor (1978) reported that patients with damage to a particular part of the right hemisphere could recognize objects when they were presented in a typical view but not when presented in an unusual view. These patients also found it very difficult to say whether two photographs (presented simultaneously) were of the same object when one image was a typical view of that object and one an unusual view.

One explanation for this effect is that the patients could not transform the two-dimensional representation of the unusual view of the object into a 3D model description. However, as well as it being difficult to establish the central axis of an object presented in an unusual view, it is also likely that rotation would cause some key features of the object to become hidden. Humphreys and Riddoch (1984) prepared images of objects in which *either* a critical feature was obscured *or* where the central axis had been foreshortened through rotation. These images were presented to patients similar to those tested by Warrington and Taylor. The patients had far more problems recognizing the axis-foreshortened objects than those with a key feature hidden. The results of these studies do offer some evidence that axis location may play a central role in generating a 3D model description of an object.

### **3.3 Biederman's theory**

Marr and Nishihara's work has been extended and adapted in several related theories of object recognition. The most influential of these was proposed by Biederman in 1987. Biederman's theory (1987a) was also based on representing complex objects using a series of more simple primitives. Unlike Marr and Nishihara, Biederman did

not restrict these primitives to generalized cones. Instead he proposed that the basic building blocks for describing an object were a set of basic shapes such as cylinders and cubes known as **geons** (an abbreviated form of the phrase ‘geometric ions’). Many of these geons are generalized cones, but they also include other 3D shapes that are very useful in representing common objects. A sub-set of geons is shown in the top part of Figure 4.19.



**Figure 4.19** A selection of geons

Source: Biederman, 1987b

Biederman suggested that approximately 36 geons are needed in order to produce descriptions of all common objects. As with Marr and Nishihara’s theory, more complex objects are represented by several different components and the division into components is based on areas of concavity.

The principal way in which Biederman’s theory diverges from Marr and Nishihara’s approach is the way in which a 3D description is formed from information in a 2D image – in other words, how the information in the primal sketch can be used to generate a 3D object-centred description. Biederman proposed that Marr’s contour generators are not necessary to recover 3D shape, as each geon will have a key feature that remains invariant across different viewpoints. Thus, all that needs to be done is to locate these key features in the 2D primal sketch. Each feature can then be matched to a geon so that a 3D structural description of the object is generated. This description is then matched against those stored in memory.

Behind the concept of key features that remain invariant across viewpoint is the idea that some regular aspects of a 3D shape will tend to remain constant in any 2D image that is formed of that object. Biederman termed these ‘nonaccidental’ properties to distinguish them from any regularity that was due simply to a particular viewpoint.

Biederman listed five nonaccidental properties:

*Curvilinearity* – a curve in the 2D image is produced by a curve on the object.

*Parallelism* – lines that are parallel in the 2D image will be parallel on the object.

*Cotermination* – two or more edges that terminate at the same point in the 2D image will terminate at the same point on the object.

*Symmetry* – if the 2D image is symmetrical then the object will contain the same axis of symmetry.

*Collinearity* – a straight line in the 2D image is caused by a straight line on the object.

Choosing which geon to use in order to represent an object (or part of an object) is then simply a matter of detecting these nonaccidental properties and selecting a geon that shares them. For example, the 2D image of a ball will be a circle and will therefore contain no parallelism, cotermination or collinearity, but will contain curvilinearity and an almost infinite degree of symmetry. The only geon to share these properties is a sphere, so the 3D shape of the ball is correctly described by a spherical geon.

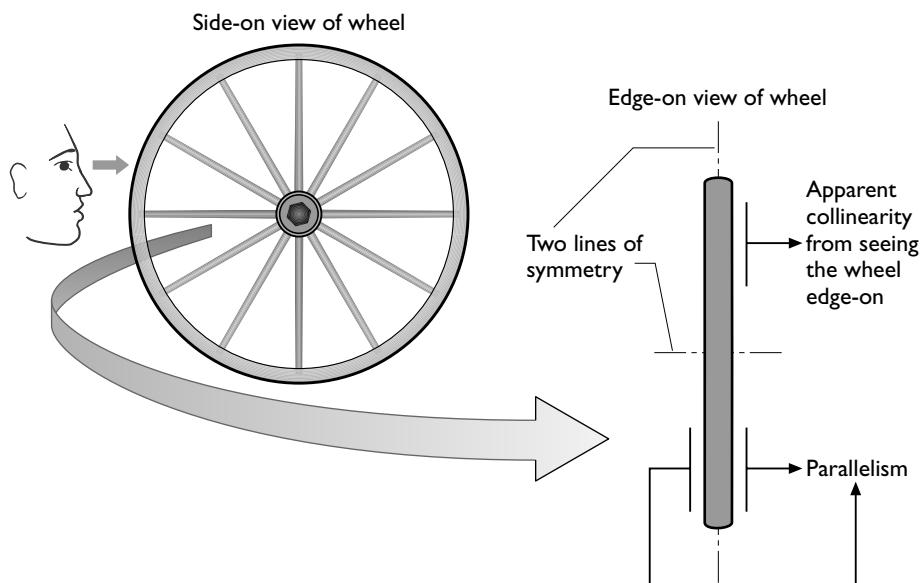
Although these assumptions allow apparently ambiguous 2D images to be turned into an accurate 3D description, they can also lead to misinterpretation. For example, if you look at the wheel of a bicycle that is directly in front of you so that the wheel is viewed edge-on, its edges will appear to have the following nonaccidental properties (see Figure 4.20):

*Collinearity* – the two vertical edges will appear as straight lines.

*Symmetry* – there will be two lines of symmetry, one horizontal and one vertical.

*Parallelism* – the two vertical edges will appear parallel.

However, the first of these nonaccidental properties (collinearity) will be incorrect as a wheel does not contain any straight edges. We only see straight edges because of the viewpoint.



**Figure 4.20** Apparent nonaccidental properties of a wheel viewed edge-on

Although describing an object using nonaccidental properties to select geons can lead to problems, there is evidence that supports Biederman's theory. The premise that concavities are used to divide the object into components (this premise was also used by Marr and Nishihara) was studied by presenting participants with images of objects that had part of their contours deleted. Deleting the part of the contour that corresponded to a concavity (that therefore occurred between components) resulted in a greater disruption to recognition than deleting part of the contour from elsewhere on the object (Biederman, 1987a).

The production of an object description that is independent of viewpoint is a crucial stage in the theories of both Marr and Nishihara and of Biederman. So is there evidence that recognition does involve the generation of an object-centred description rather than relying purely on viewer-centred descriptions?

To investigate the extent to which recognition is object-centred, Biederman and Gerhardstein (1993) used a technique known as repetition priming, where the presentation of one stimulus will make recognition of a related stimulus faster and/or more accurate. The idea behind their experiment was that if an object-centred description were being formed, then presenting one particular viewpoint of an object should facilitate (or prime) recognition of the same object presented in a different view. Their results showed that one viewpoint of an object did prime recognition of a separate viewpoint, as long as the change in the angle of viewpoint was not more than 135 degrees. However, even if the viewpoints were less than 135 degrees apart, if one or more geon was hidden between the first and second view, then the amount of priming was reduced. This result supports both the idea that an object-centred description is generated (otherwise different viewpoints should not prime each other), and that this makes use of geons.

However, other researchers have reported results that do not appear consistent with Biederman and Gerhardstein's findings. Bulthoff and Edelman (1992) found that participants were generally unable to recognize complex objects that were presented in a novel viewpoint, even if the view of the object was one that should have allowed the generation of an object-centred description. In the end, it is unlikely that recognition is *completely* reliant upon the generation of object-centred descriptions such as those suggested by Marr and Nishihara (1978) and by Biederman (1987a), and there may well be tasks that involve viewpoint-dependent recognition (Tarr, 1995).

One task that it is hard to incorporate into either Marr and Nishihara's or Biederman's theory is that of within-category discrimination. By representing objects as models consisting of either generalized cones or geons, a wealth of information is inevitably lost. For example, it is very likely that two collie-shaped canines would be represented as identical 3D models, yet it is possible to tell a border collie from a rough collie and even to tell specific dogs apart.

It makes sense that there should be more than a single way of arriving at such a complex cognitive achievement as object recognition. In the theories we have examined in this section, the process of recognition has been conceived of as almost wholly passive and based on a single retinal 'snapshot' or view. As we have stated previously, there are different types of recognition and different ways of achieving it, including taking a more active approach.

## Summary of Section 3

---

- Objects can be recognized from many different angles, suggesting that the process of recognition may be based on the generation of a 3D object-centred description.
- Marr and Nishihara (1978) suggested a theory of object recognition based on generating 3D models. This was achieved by: deriving the shape of an object from the 2½D sketch; dividing it into ‘primitives’ using areas of sharp concavity; generating an axis for each of these components; and representing each component via a generalized cone.
- The 3D models were hierarchical in nature, and so include both global and detailed information stored in a hierarchically organized catalogue.
- Biederman (1987) suggested a similar theory based on using the nonaccidental properties of an object to generate a description in terms of a series of basic volumetric forms known as geons.
- Although there is evidence that supports the approach taken by Marr and Nishihara and by Biederman, there are some forms of recognition which are difficult to explain using their theories.

## 4 Face recognition

Another type of recognition, and one that is very problematic for the 3D model approaches we have looked at so far, is that of recognizing faces. If we return to Humphreys and Bruce’s model of object recognition shown in Figure 4.2, we can see that these theories have concentrated on the ‘perceptual classification’ stage of the process. Although this stage may provide information useful for navigation and basic interaction with the objects we find in the environment, more complex interaction is often necessary. For example, when you are confronted by a person, you want to know not only that there is a human face in front of you, but *whose* face it is. This requires a much finer level of distinction than simply recognizing a sphere as a sphere; you must be able to tell which *specific* face is in front of you. As we shall see in Sections 4 to 7 of this chapter, the need to recognize individual faces has led to theories and research concentrating on different issues from that conducted within the area of more general object recognition.

Faces can be categorized at several different levels. At one level, we decide that the stimulus is a face as opposed to some other object. At another level, we decide that the face is female or male or derive other semantic information such as ethnic origin. We may even make attractiveness judgements. Importantly, we also decide whether the face is familiar or unfamiliar. If the face is familiar, there is also the need to decide to whom the face belongs and it is at this level that faces are rather different from other objects. It is this within-category judgement, which is like recognizing a specific cat or a specific cup, that sets face recognition apart from

object recognition more generally and is regarded as more visually demanding because the differences between faces can be fairly minimal.

Tanaka (2001) has found evidence to liken this level of face recognition to expert recognition – for example, the expertise that certain individuals acquire through training at bird-watching or x-ray analysis. But whereas only some specifically trained people achieve object expertise, face expertise is a general expertise that we all share and acquire without specific training. Whether or not this face expertise is the result of an innate processing system or the expression of a learned skill is a matter of debate and an issue we will return to in Section 7 of this chapter.

## 4.1 Recognizing familiar and unfamiliar faces

So how good are we at recognizing faces and identifying people? You already saw in Activity 4.2 that it was possible to recognize a face that was familiar to you (Paul McCartney) despite quite large changes in appearance. In fact, when you think about it, you are able to recognize your family and friends from any angle, under different lighting conditions and even when they age or change their hairstyle, and you are still likely to be able to do this in 30 years time. There is evidence to suggest that we can remember the names and faces of school-friends over long periods of time; recognition tests revealed hardly any forgetting over a 35-year period (Bahrick *et al.*, 1975). This is not the case with all the faces we encounter though. Later work by Bahrick investigated the ability of college teachers to recognize former students taught over a 10-week period (Bahrick, 1984). The teachers had met these students three to five times a week. Although the level of correct face recognition for those taught recently was reasonably high at 69 per cent, this dropped as the number of intervening years increased so that after 8 years only 26 per cent of the former students were correctly recognized.

What about faces that are not so familiar and that we've only seen once? A number of face-learning experiments have been conducted (e.g. Yin, 1969) and these have found that, when given an immediate recognition test, participants performed extremely well. (For example, Yin observed that participants correctly recognized 93 per cent of the faces previously shown to them). However, if the picture of the face shown in the recognition test depicted a different viewpoint or expression, then recognition rates dropped (e.g. Bruce, 1982), suggesting that what is being tested is 'recognition of a specific picture of a face' rather than 'face recognition' as we encounter it in everyday life.

Indeed, as you will see in Box 4.1, research has demonstrated that unfamiliar face recognition appears to be quite different from familiar face recognition.

**4.1****Research study****Recognizing unfamiliar faces in matching tasks**

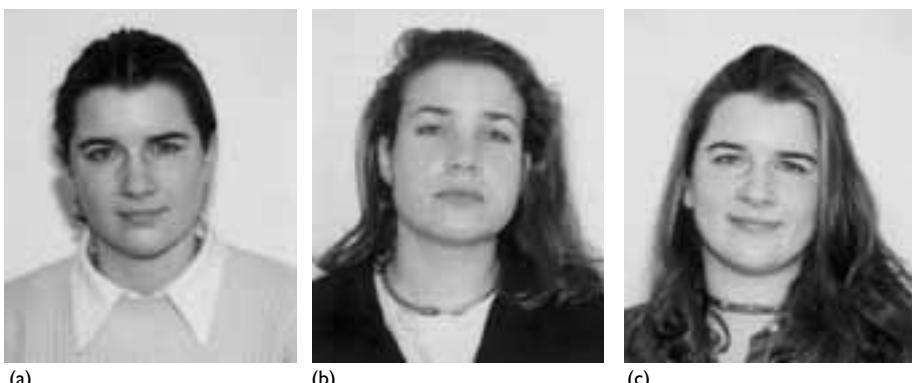
Even matching unfamiliar faces that are presented simultaneously (a task that does not test our memory) appears to be surprisingly difficult. In a field experiment, Kemp *et al.* (1997) looked at how well cashiers could match shoppers to credit cards bearing their photographs. They found that cashiers would frequently accept credit cards depicting a photograph of someone who bore a resemblance to the shopper (the correct decision rate to reject the card was only 36 per cent). Even when the photograph was of someone who bore no particular resemblance to the shopper but was of the same sex and ethnic background, the correct decision rate to reject the card was only 66 per cent (see Activity 4.1).

Other studies have demonstrated that we are not very good at matching two similar high quality photographic images when the face is unfamiliar. Bruce *et al.* (1999) showed participants a high quality video still of an unfamiliar young male target which was then presented in a line-up of similar images of nine other young men. Even when told that the target was definitely present in the line-up, participants picked it out accurately on only 80 per cent of the trials. If not told that the target was present, or if the pose of the target was varied between initial presentation and test, then performance was still worse. In fact the performance of these participants has been matched or even exceeded by that of an automatic face recognition system tested on the same images (Burton *et al.*, 2001).

Of interest too are the findings of a study looking at our ability to recognize unfamiliar faces by touch. Kilgour and Lederman (2002) found that when participants explored the faces both visually and tactually, performance was no better than when the faces were explored by touch alone.

**ACTIVITY 4.4**

Look at the images of three faces shown in Figure 4.21. Which of the images to the left (a or b) do you think is of the same woman as that in the right-hand image (c)? These images are examples of images that were used on photo-credit cards in the study conducted by Kemp *et al.* (1997).



**Figure 4.21** Three faces

### COMMENT

The correct answer is that the left-hand image (a) is of the same woman shown in (c), but cashiers often refused to accept it due to the change in hairstyle. However, the image in the centre (b) was often incorrectly accepted as being of the woman to the right (c).

---

We will focus the rest of our discussion of face recognition largely on our ability to identify familiar faces and will start our discussion by considering some of the errors people make. These errors provide us with important information about the different systems and processes that may be involved in face recognition. Importantly, models of face recognition need to be able to account for such errors.

## Summary of Section 4

---

- Face recognition is an example of a within-category judgement task.
- Our ability to identify familiar faces is extremely good and relatively unaffected by pose, lighting or viewpoint.
- Recognition of unfamiliar faces is much poorer and is influenced by changes in pose, lighting or viewpoint.

## 5 Modelling in face recognition

The theories of object recognition we have looked at previously centred on matching the description of an object that is in view with a stored representation. Although face recognition also involves similar matching processing, this is not usually considered the end point. In addition to matching the face we also need to access relevant semantic information and, preferably, the person's name.

### ACTIVITY 4.5

---

Although we may have face expertise, we do make mistakes. Before reading on, reflect for a moment and recall the last time you discovered that you failed to recognize someone you know or you mistakenly thought you recognized someone you didn't know.

---

In a diary study, Young *et al.* (1985) asked 22 participants to make a record of the mistakes they made in recognizing people over an eight-week period. The recorded errors or difficulties tended to fall into different categories as shown in Table 4.2.

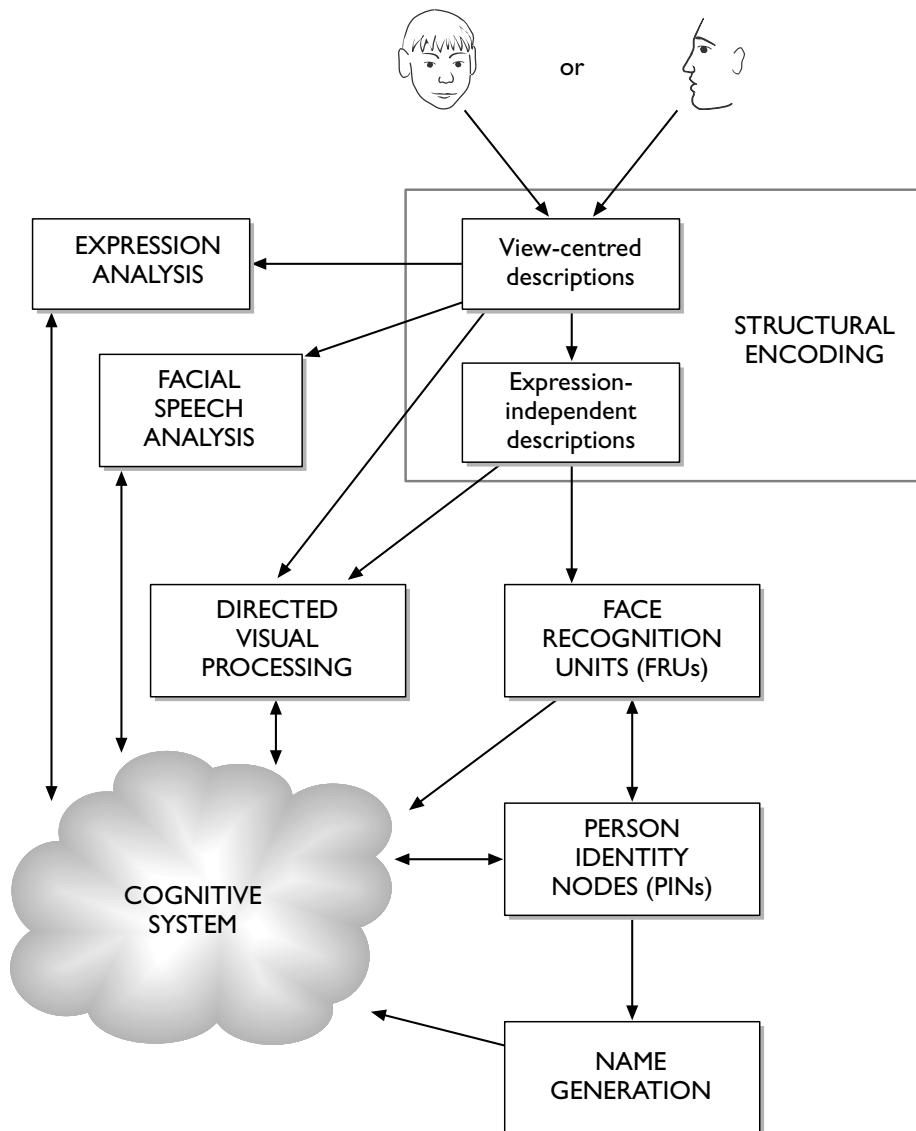
**Table 4.2** The main types of everyday errors in face recognition revealed by Young *et al.* (1985)

Types of everyday errors	Number of errors
Person misidentified	314
Person unrecognized	114
Person seemed familiar only	223
Difficulty in retrieving full details of the person	190
Decision problems	35

What do these different categories mean? '*Person misidentified*' refers to those occasions when someone unfamiliar is misidentified as someone familiar and '*Person unrecognized*' refers to occasions when someone familiar was thought to be someone unfamiliar. Both may arise because of poor viewing conditions (i.e. it is a bit dark) or because we know the person only slightly. '*Person seemed familiar only*' refers to those occasions when you recognize someone as being familiar but no other information comes to mind immediately, and '*Difficulty in retrieving full details of the person*' refers to occasions when only some semantic information is retrieved and not, for example, their name. These errors often occur when the familiar person is seen outside the context in which they are usually encountered. Finally, '*Decision problems*' refer to those occasions where you think you recognize the person but decide it cannot be them, perhaps because you believe they are currently in another country.

The pattern of these errors suggests that, although we might retrieve previously learned semantic information about a person without recalling their name, we will never recall their name without also retrieving relevant semantic information. However, before we can recall either semantic information or a name, we must realize the face is familiar.

These findings on everyday errors are consistent with the notion that the recognition of faces involves a sequence of processes using different types of information. Hay and Young (1982), Young *et al.* (1985) and then Bruce and Young (1986) refined a cognitive theoretical framework or model of person recognition involving such a sequence of stages. On meeting people, we encode their faces. This encoded information may activate **face recognition units** (FRUs) that contain stored information about the faces we are familiar with. If there is a reasonable match between what has been encoded and what is stored in the recognition unit, then the recognition unit will be activated and allow access to semantic information about the person's identity, such as their occupation, stored in **person identity nodes** (PINs). It is only once the PIN for a face has been activated that their name can be generated. A **cognitive system** is also involved, as the information provided by the recognition system must be evaluated. As the diary study above indicated, errors in face recognition can arise because of decision problems. For example, if we know that the person doesn't live or work nearby, that knowledge may override what our recognition system is telling us and hence we may doubt that we have correctly identified the person.



**Figure 4.22** Bruce and Young's functional model for face recognition

Source: Bruce and Young, 1986, p.312

The Bruce and Young (1986) functional model for face recognition is presented in Figure 4.22. As you can see, there are separate routes for facial expression analysis, facial speech analysis, and face recognition; and face recognition progresses through a sequence of stages from FRUs to PINs to name generation.

The notion that different types of information are sequentially accessed is also supported by the results of experiments conducted in the laboratory. For example, Hay *et al.* (1991) showed participants 190 famous and unfamiliar facial images and asked them to decide whether or not each face was familiar and to state the person's occupation and the person's name. Participants did not retrieve a name without also being able to name the occupation, thus supporting the notion that semantic 'person identity' information is retrieved *before* the person's name. Other studies (e.g. Johnston and Bruce, 1990), looking at how quickly we can complete a particular task, have shown that faces can be classified as familiar more quickly than they can be classified by occupation, and furthermore that classifications that require accessing the person's name take longer than classifications involving a person's occupation or other semantic properties. These findings support the notion that perceptual classification, judging the familiarity of a person, takes place *before* semantic classification and that a person's name is accessed last. They also provide a nice demonstration of how the findings from the laboratory may support those derived in a more everyday study of face recognition, such as Young *et al.*'s (1985) diary study.

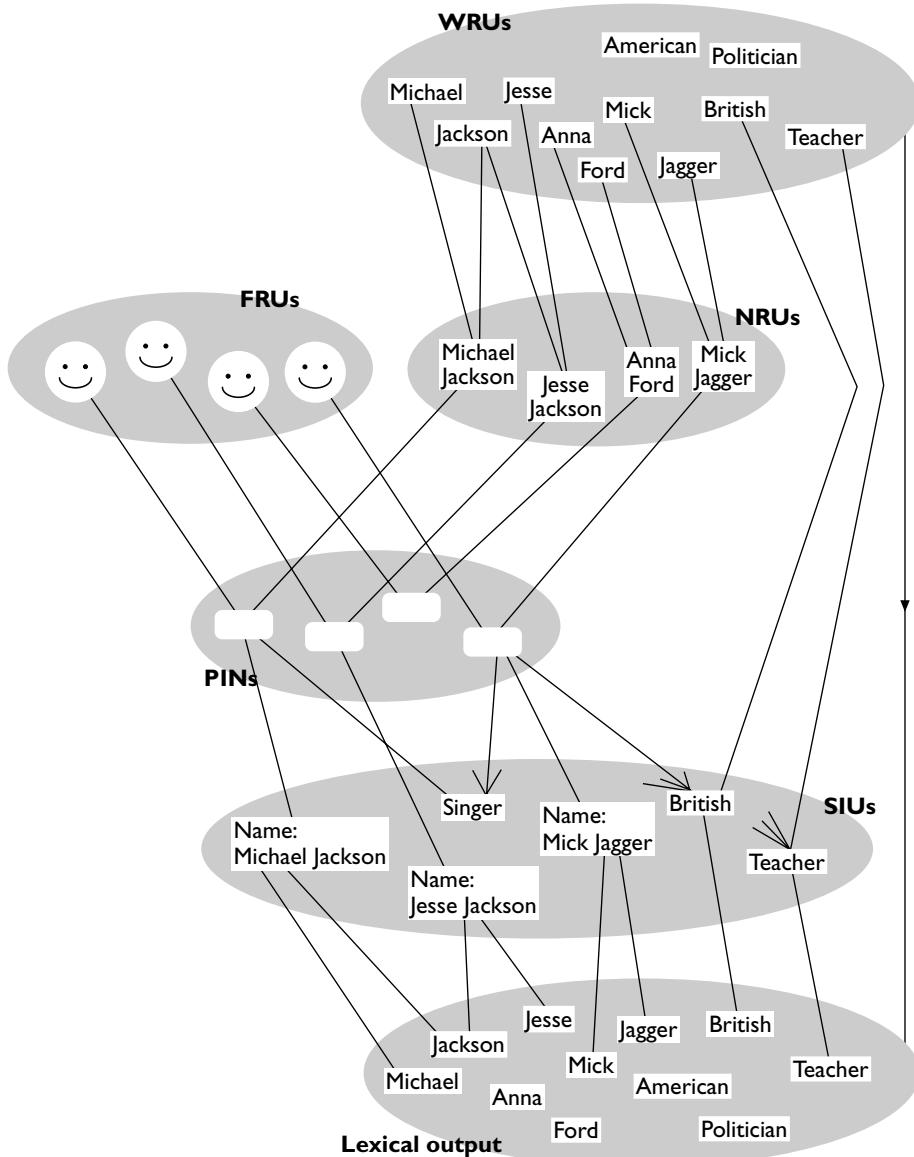
## 5.1 A connectionist model of face recognition

The **IAC model** (e.g. Burton *et al.*, 1990; Burton and Bruce, 1993) is a connectionist model (recall the discussion of connectionism in Chapter 1) of face recognition and an extension and implementation of the Bruce and Young model described above. IAC stands for 'interactive activation and competition network'. As this model is a computer simulation of face recognition it has been tested by seeing how compatible it is with the available evidence, and by looking at the predictions it generates.

The model comprises units which are organized into pools (see Figure 4.23). These pools contain:

- *FRUs (face recognition units)*: For every familiar person, there is one FRU in the model. These are view-independent and seeing any recognizable view of a face will activate the appropriate FRU. These representations allow perceptual information to be mapped onto stored memories. (This is basically what was suggested in the Bruce and Young model.)
- *PINs (person identity nodes)*: This is where a face is classified as belonging to a person, and there is one unit per known person.
- *SIUs (semantic information units)*: Relevant semantic information is stored here, e.g. occupational category.
- *Lexical output*: Units representing output as either words or name.

The IAC model also includes a route based on word recognition. The pool of WRUs (word recognition units) represents an input lexicon containing both specific names and more general information, such as nationality or occupation. Words which are names have direct links to a pool of NRUs (name recognition units), which are linked to PINs in the same way as FRUs. The WRUs which do not correspond to names are linked to SIUs.



**Figure 4.23** The central architecture of the IAC model

Figure 4.23 shows how the pools are connected. The input systems (FRUs and NRUs) join to a common set of person identity nodes (PINs) and these are linked to units containing semantic information (SIUs). Each of the pools is illustrated here with just a few examples of the units they might contain. Many SIUs will be shared and here many people will be represented with such information as 'teacher' or 'British'.

Recognizing a face is modelled in the following way: seeing a face will activate an FRU which in turn increases activation in the relevant PIN. As PINs

are linked to SIUs, activation of the PIN will bring about activation in the relevant SIU. The notion that different types of information are sequentially accessed is therefore still present in this connectionist model. If a certain threshold is achieved in the PIN, then this signals familiarity. An important point to note is that different types of information come together at the PIN stage, including information from recognition systems specialized for faces as well as those specialized for the recognition of written or spoken names, and familiarity is judged on this pooled information.

We mentioned before that IAC stood for ‘interactive activation and competition network’. The ‘interactive activation’ arises from the links between units in different pools which are *excitatory*: the FRU for Mick Jagger’s face excites or activates the PIN for Mick Jagger which in turn excites semantic information units for the name ‘Mick Jagger’, the occupation ‘singer’ and the nationality ‘British’. These excitatory links are bidirectional so that excitation also runs in the opposite direction from ‘singer’ to Mick Jagger’s PIN and Mick Jagger’s FRU. However, within each pool, links between units are *inhibitory* (these links are not shown in Figure 4.23), so this is where ‘competition’ arises. Excitement in the FRU for Mick Jagger will inhibit activity in the other FRUs, just as excitement in Mick Jagger’s PIN will inhibit activity in the other PINs and excitement in one SIU will inhibit activity in another SIU. But, the SIU for Mick Jagger (which might be ‘singer’) will also excite many other PINs (in this example, those belonging to all other singers). This means that activation of PINs will not be limited solely to the specific person in question but that some activation will also occur for anyone who is semantically related (e.g. shares the same occupation). Thus, the model incorporates the results of experiments that have shown priming effects – that you are quicker to recognize Bill Wyman if you have already seen Mick Jagger. Generally, the strength of this connectionist model is that it can account for findings from laboratory studies as well as for the everyday errors described by Young *et al.* (1985).

## Summary of Section 5

---

- Everyday errors suggest that recognizing faces involves sequential access to different types of information.
- A cognitive model of person recognition involving such an idealized sequence of stages has been developed (Bruce and Young, 1986).
- IAC is a connectionist model of face recognition which is an extension and implementation of this cognitive model.

## 6 Neuropsychological evidence

**Prosopagnosia**, the inability to recognize faces whilst maintaining the ability to recognize other objects, is a well-documented phenomenon. However, cases of 'pure' prosopagnosia are exceptionally rare. It is more common to see deficits affecting other visual categories too. The recognition of all familiar faces is affected, regardless of their semantic categories (so it is not the case that the failure to recognize a face is restricted to faces of celebrities or politicians). However, as recognition from other cues, for example voice, usually remains unaffected, the condition is specific to visual recognition of faces and is not a more general impairment of the recognition of personal identity. Also, the ability to distinguish between faces is often preserved.

In this section, we shall focus on two key findings that have emerged from investigations of prosopagnosia: first, that identification of expression appears to be independent from face identification; and, second, that face recognition and awareness of face recognition might also be independent of one another. It is possible that although prosopagnosics are unable to recognize faces consciously or overtly, certain types of nonconscious response may be preserved. We shall examine how the IAC model may account for this.

As mentioned in Section 5, models of face recognition have proposed a route for face identification that is independent of emotional expression, and this independence has received support from experimental work and from neuropsychological research. In many cases of prosopagnosia, the ability to recognize facial expressions may be unaffected. Young *et al.* (1993) looked at ex-servicemen with unilateral brain injuries and tested familiar face recognition, unfamiliar face matching and analysis of emotional facial expressions. Analysis of accuracy data showed evidence of selective impairments in each of these three abilities. For example, one participant with a right hemisphere lesion was selectively impaired in identifying familiar faces, whereas a different participant, also with a right hemisphere lesion had problems only with matching unfamiliar faces. A number of other participants with left hemisphere lesions were only impaired on the facial expression tasks. Response latency data also supported the notion of a selective deficit of facial expression processing but suggested that impairments of familiar face recognition and unfamiliar face matching were not entirely independent from one another. The findings from this study thus provide strong support for the notion that facial expression analysis and face identification seem to proceed independently of each other (and also some support for the notion that the ability to recognize familiar faces and to match unfamiliar faces may be selectively and independently impaired).

Previously, when describing models of face recognition, we did not draw a distinction between face recognition and awareness of recognition. However, neuropsychological research on prosopagnosia suggests that the distinction is important. Bauer (1984) monitored changes in autonomic nervous system activity via changes in **skin conductance response (SCR)**. These changes signal an affective or emotional reaction (you may remember reading in Chapter 2 on attention how a closely related response, GSR, was measured to look at

unconscious processes). Bauer showed LF, a participant with prosopagnosia, a face and read out a list of five names, whilst simultaneously measuring SCR. If LF was asked to pick the correct name, he performed no better than at chance. In other words, LF was overtly unable to recognize familiar people from their faces. However, LF showed a greater SCR when the correct name was read aloud compared with the incorrect names. Thus, LF was showing an affective or emotional response, but this response was not a conscious one. The term **covert recognition** is used to describe this nonconscious recognition or emotional response to the faces.

Since Bauer's work, many studies have investigated covert recognition and the issue is not whether this type of face recognition exists but how to interpret it. Bauer proposed that separate neural pathways are responsible for two independent routes to recognition, one for conscious overt recognition and one for nonconscious covert recognition. Although questions remain over exactly how overt and covert recognition processes are mediated and how these processes normally become integrated, there is support for the involvement of the two major neural pathways (see Box 4.2).

## 4.2

## Research study

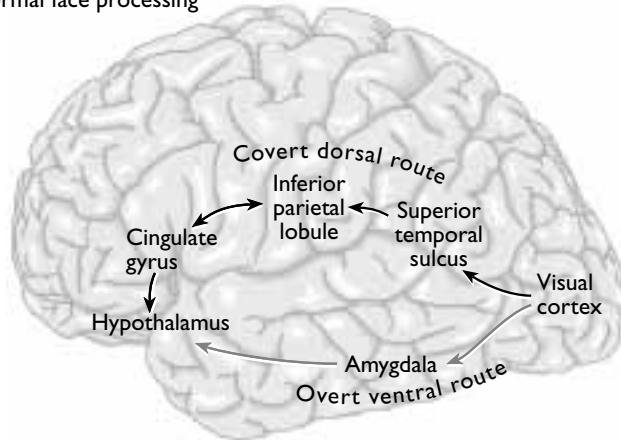
### Capgras delusion

Capgras delusion usually occurs as part of a psychiatric illness although it can result from brain injury. A person with Capgras delusion believes firmly that someone they know, usually a relative or close friend, has been replaced by an impostor, double, robot or alien. Sometimes the delusion relates to objects; for example, the sufferer may believe that tools, ornaments or other household objects have been replaced by doubles. Face and object Capgras delusion do not usually co-exist, and the disorder tends to be specific to one domain. The key point here is that individuals with a face Capgras delusion recognize a face but simultaneously refute its authenticity. Exactly why those with Capgras delusion adhere to the belief that the person must be an impostor is still being debated.

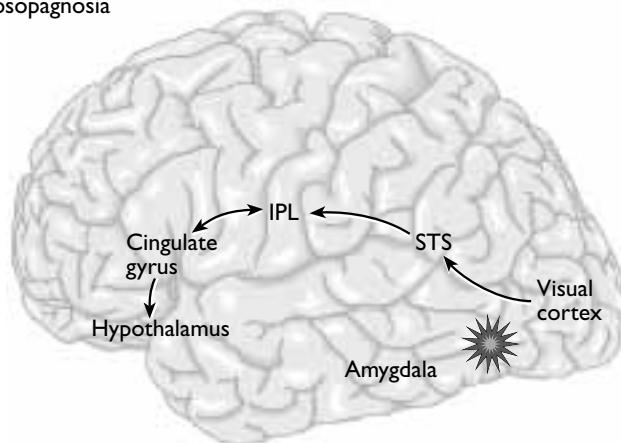
Ellis and Young (1990) suggested that Capgras delusion may be a 'mirror image' of the impairments underlying prosopagnosia. Bauer (1984) proposed that the neuroanatomical pathway involved in overt recognition was the 'ventral visual-limbic pathway' whereas the pathway involved in covert recognition was the 'dorsal visual-limbic pathway'. Ellis and Young suggested that the Capgras delusion resulted from damage to such a dorsal route, so that sufferers would recognize the familiar person but not receive supporting affective information. Their prediction that individuals with Capgras delusion would recognize familiar faces but would fail to show an autonomic emotional response to these familiar faces has received support from several studies (e.g. Hirstein and Ramachandran, 1997). Whilst overt recognition is intact, covert recognition seems to be impaired.



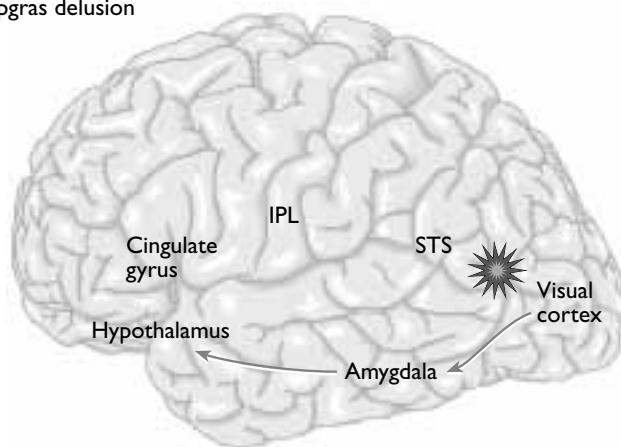
(a) Normal face processing



(b) Prosopagnosia



(c) Capgras delusion



**Figure 4.24** The dorsal and ventral routes in normal face processing (a), prosopagnosia (b) and Capgras delusion (c)

Source: Ellis and Lewis, 2001, Figure 3, p.154



Figure 4.24 shows normal face processing (a), with a darker arrow showing the covert dorsal route and a lighter arrow the overt ventral route. In prosopagnosia (b) the overt ventral route is thought to be damaged, and in Capgras delusion (c) the covert dorsal route is thought to be damaged.

A different issue is whether those individuals who retain covert recognition can be helped to overcome their disorder. Could covert recognition be turned into overt recognition? Sergent and Poncet (1990) were the first to demonstrate such provoked overt recognition. In their study, PV was shown eight faces of famous people from the same semantic category and she was unable to identify them. However, when she was told that they all had the same occupation and she looked at the faces again, she was able to say that they were all politicians, name seven of the people, and recall biographical information about the eighth person. This and other later studies (e.g. Diamond *et al.*, 1994) have shown that provoked overt recognition can occur under certain experimental conditions, and this provides some hope for rehabilitative work.

Can the IAC model accommodate the pattern of deficits described here? Covert without overt recognition is explained in terms of attenuation (or weakening) in the connections between the FRUs and PINs. This means that when a face is seen, and the FRU is activated, the weakened FRU-PIN connection strength means that excitation of the corresponding PIN is not raised above the threshold for the face to be recognized overtly. However, this weakened activation may be sufficient to raise the excitation of the PIN above its resting level, mediating covert recognition. Provoked overt recognition is explained in the following way. Telling PV that the faces are related is equivalent to strengthening the PIN-SIU connections. Unlike FRU-PIN connections, PIN-SIU links are assumed to remain intact in instances of prosopagnosia where covert recognition is observed. Once these connections are strengthened, activation is passed back from the shared SIUs to the relevant PINs. These then achieve threshold and the faces are recognized overtly. Simulations with the model confirmed this particular prediction – provoked overt recognition was successfully modelled (Morrison *et al.*, 2001).

So, as a model of face recognition, the IAC model is impressive in that it can account for a wide range of data from studies on face recognition. Whilst there are other models of face recognition, some of these are based on a narrower range of evidence; for example, they may have sought only to account for the findings from neuropsychological studies. As we have seen here, IAC is compatible with everyday, laboratory and neuropsychological findings.

## Summary of Section 6

---

- Prosopagnosia is the inability to recognize faces although expressions and other objects may still be correctly identified.
- Covert face recognition, shown by autonomic responses to faces, may however be spared.

- Overt conscious face recognition and covert nonconscious face recognition are different types of face recognition that may be mediated by different neural pathways.
- Capgras delusion may be a mirror image of prosopagnosia in terms of which system remains intact and which system is damaged.
- Provoked overt recognition has been achieved in some studies and has been successfully modelled using the IAC model.

## 7 Are faces ‘special’?

In this last section we return to the issue of the difference between face recognition and object recognition, and in particular to face expertise and how we are able to discriminate so readily between faces. There are several important issues that the literature has addressed:

- 1 Is there a neuroanatomical location that underlies face processing and, if so, does this mean that face processing is unique and qualitatively different from the processing of other types of visual stimuli?
- 2 Is face processing an innate or learned skill? Have we developed a face expertise because of constant exposure to faces and practice at differentiating between them or is there an innate ability?
- 3 How important are the individual features of the face, the relationships between the features, or the three-dimensional structure? Do we process the individual facial features or the face as a whole?

In the last section, we looked at the syndrome of prosopagnosia and found that research implicated several neurological pathways. Of particular interest is that prosopagnosia can leave object recognition relatively intact and, in turn, face recognition has been spared in cases where object recognition has been impaired (a double dissociation). Studies using the technique of functional magnetic resonance imaging (fMRI) have found facial stimuli to activate an area in the fusiform gyrus in the posterior temporal lobes (especially in the right hemisphere) whilst nonface objects activated a different area. There is also the observation of cells specialized for faces within the monkey temporal lobe – these cells respond selectively to faces of humans and/or monkeys but not to other stimuli (e.g. geometrical shapes and bananas). There is, therefore, evidence to suggest that the processing of faces is mediated by specific areas of the brain, that there is cortical specialization for faces. But does this mean that face recognition is *unique*, that the processes used for recognizing faces are qualitatively different from those used for recognizing other visual stimuli?

There is support for the notion that there is a special mechanism from birth for processing facial information, as newborn babies show a preference for face-like visual patterns. Rather than an innate neural mechanism that processes faces, Johnson and Morton (1991) suggested that there is a mechanism that makes newborns attentive to faces, and this innate attentional bias then ensures that any system for learning visual stimuli receives a lot of face input and learns about the

individual characteristics of faces. Although there is a ‘kick-start’ mechanism which gives face processing in newborns a special status, this serves to guide subsequent learning and soon other processing systems will come into play (these may or may not be unique to faces).

One reason to think that face recognition is a special type of recognition, distinct from other object recognition, is that faces all tend to look alike in that they have similar features in similar positions. Given this similarity, it could be that we have to make use of a different form of visual information to recognize a face from that used to recognize, for example, a table. Some evidence that this is indeed the case comes from studies that have demonstrated that inverting, or turning upside down, visual stimuli disproportionately impairs our ability to recognize faces compared with our ability to recognize objects. This is known as the **inversion effect**. Yin (1969) and other studies since (e.g. Johnston *et al.*, 1992) have shown that inverting a photograph of a face disrupts recognition more than does inverting an image of an object. Yin looked at the influence of inversion on faces and other stimulus material including houses and aeroplanes. Although recognition memory was better for upright faces than for other material, when the stimuli were turned upside down, recognition for faces was worse than that for other material. The key question is whether this peculiar reversal of recognition accuracy for faces (from best upright to worst inverted) supports the notion that faces are processed differently from other stimuli or whether there is an alternative explanation.

Diamond and Carey (1986) investigated an alternative hypothesis, namely that the effect of inversion on faces was a result of our perceptual mechanisms becoming ‘tuned’ to seeing this special type of visual stimulus in the usual upright orientation. This ‘tuning’ or expertise would then be ‘lost’ when we see them inverted. Their research considered whether the inversion effect was indeed specific to human faces or whether it would in fact arise when using any class of visual stimulus with which we have a large amount of experience. To investigate this, Diamond and Carey selected participants to include both people who were not interested in dogs and people who were dog experts (mainly dog-show judges, breeders/handlers or people with a sustained interest in dogs). These participants were shown photographs of both human faces and dogs (body profiles – see Figure 4.25) and told to look at each photograph and try to remember it. Analysis revealed that whereas all participants recognized upright faces better than inverted faces, dog experts also recognized upright dogs better than inverted dogs. This finding has been interpreted as supporting the notion that the inversion effect is acquired as a result of expertise and is not a ‘face-specific’ effect.

What changes then in the way we process faces as we acquire this expertise? Diamond and Carey proposed a distinction between first-order and second-order relational properties. *First-order relational properties* refer to the spatial relationships among parts of the face; for example, the eyes are above the nose and the mouth below the nose. Faces cannot be distinguished according to their first-order relational properties as they all share the same basic arrangement or configuration. However, first-order relational properties help us detect that a visual stimulus is a face – a necessary step before identifying the face. *Second-order relational properties* refer to the differences in this basic configuration. This refers to the differences in the way the features are arranged in relation to each other; for example,



(a) Inspection items that participants were asked to remember



(b) Recognition items: participants were asked to judge which of the stimulus items they had seen previously

**Figure 4.25** Examples of the dog stimuli used by Diamond and Carey

Source: Diamond and Carey, 1986, p.112

wide-set eyes with a low forehead versus narrow-set eyes and a high forehead. Expertise results in a greater sensitivity to these second-order relational properties, as it is these properties that individuate members of the same class, such as human faces.

There is support for the notion that inversion influences our sensitivity to second-order relational properties. For example, Searcy and Bartlett (1996) presented participants with photographs of grotesque looking faces. They created images where they had either distorted individual facial features (eyes and mouths) or they had distorted the spatial relations between the features (see Figure 4.26). They then presented these manipulated images in upright and inverted orientations. Participants rated the grotesqueness of the images and results showed that images of faces with distortions to the spatial relations between the features were rated as



**Figure 4.26** Examples of stimuli used by Searcy and Bartlett (1996): the pair labelled 'A' shows a normal image and one with distorted facial features; the pair labelled 'B' shows a normal image and one with spatial distortion

Source: Searcy and Bartlett, 1986, Figure 1, p.907

less grotesque when presented inverted rather than upright; inversion failed to reduce ratings of grotesqueness when the distortions were performed on the features. These findings support the notion that inversion disrupts our processing of spatial relationships between the features.

Research like this suggests our expertise in (upright) face recognition stems from the way in which these upright faces are processed as ‘configurations’, rather than as an assemblage of independent features. The term **configural processing** has been used, although this has been interpreted in a number of ways: to refer to the spatial relationships between features (i.e. second-order relational properties); to refer to the way facial features interact with one another (i.e. the way the shape of the mouth influences how the shape of the nose is perceived); to refer to holistic processing of the face (i.e. the face is perceived as a whole face pattern and not broken down into separate features); or even to refer to the basic arrangement of the facial features (i.e. first-order relational properties).

A considerable amount of research has been devoted to investigating the relative importance of this type of processing as compared with the processing of the facial features (known as featural processing or piecemeal processing). Although it is not always clear what different researchers mean by the term ‘configural’, there is agreement that configural information plays an important role in the perception and representation of upright faces. The suggestion that this reliance on configural processing is the result of learning to recognize lots of faces, and hence the result of expertise, does not rule out any input from an innate mechanism, which may have ‘kick-started’ this learning by biasing attention towards faces. However, it does not suggest that face perception and recognition involve *unique* processes which are qualitatively different from those used to process other types of stimuli. Finally, it is worth noting that research has yet to clarify the different processes involved in recognizing familiar faces as opposed to unfamiliar faces, or fully to specify the overlap between the processes involved in face identification and those used in object recognition.

In sum, although there is physiological and neuropsychological evidence supporting the existence of areas specialized for processing faces, and although there is evidence suggesting an innate ability to pay attention to faces, the processes involved in face recognition do not appear to be unique.

## Summary of Section 7

---

- Neuropsychological and physiological evidence suggests that there are specific areas of the brain that mediate face processing.
- Research on newborn babies suggest an innate ability to attend to faces.
- The inversion effect appears to be linked to our expertise in processing upright faces using configural information.
- We may develop expertise at distinguishing members of other categories of visual stimuli that also involves configural processing.
- Evidence does not suggest that the processes involved in the perception and recognition of faces are unique.

## 8 Conclusion

In this chapter we have explored different types of recognition and looked at some of the mechanisms that allow us to recognize objects and faces. In reading about recognition, you may well have got the idea that cognitive psychologists still have a lot to learn about how object and face recognition may occur. This is undoubtedly the case and a great deal of research is still being conducted in order to provide a more comprehensive and detailed theory of the cognition involved in recognition. Just as there are different types of recognition, there are also different ways of recognizing faces and objects – for example, visually or by touch – and these different ways may involve different processes. So, rather than see the theories discussed here as providing a final answer, the best way to view them is as taking some of the initial steps in this complex but interesting field.

## Further reading

- Bruce, V., Green, P.R. and Georgeson, M.A. (2003) *Visual Perception: Physiology, Psychology and Ecology*, Hove, Psychology Press.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman & Company.
- Rakover, S.S. and Cahlon, B. (2001) *Face Recognition: Cognitive and Computational Processes* (Advances in Consciousness Research), Philadelphia, PA, John Benjamins Publishing Co.

## References

- Bahrick, H.P. (1984) ‘Memory for people’, in Harris, J.E. and Morris, P.E. (eds) *Everyday Memory, Actions and Absent-mindedness*, Academic Press, London.
- Bahrick, H.P., Bahrick, P.O. and Wittlinger, R.P. (1975) ‘Fifty years of memory for names and faces: a cross-sectional approach’, *Journal of Experimental Psychology: General*, vol.104, pp.54–75.
- Bauer, R.M. (1984) ‘Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test’, *Neuropsychologia*, vol.22, pp.457–69.
- Biederman, I. (1987a) ‘Recognition by components: a theory of human image understanding’, *Psychological Review*, vol.94, pp.115–47.
- Biederman, I. (1987b) ‘Matching image edges to object memory’, *Proceedings of the First International Conference on Computer Vision*, IEEE Computer Society, pp.384–92.
- Biederman I. and Gerhardstein, P.C. (1993) ‘Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance’, *Journal of Experimental Psychology: Human Perception and Performance*, vol.19, pp.1162–82.
- Bruce, V. (1982) ‘Changing faces: visual and non-visual coding processes in face recognition’, *British Journal of Psychology*, vol.73, pp.105–16.

- Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M., and Miller, P. (1999) 'Verification of face identities from images captured on video', *Journal of Experimental Psychology: Applied*, vol.5, pp.339–60.
- Bulthoff, H.H. and Edelman, S. (1992) 'Psychophysical support for a two-dimensional view interpolation theory of object recognition', *Proceedings of the National Academy of Sciences of the USA*, vol.89, pp.60–4.
- Burton, A.M. and Bruce, V. (1993) 'Naming faces and naming names: exploring an interactive activation model of person recognition', *Memory*, vol.1, pp.457–80.
- Burton, A.M., Bruce, V. and Johnston, R.A. (1990) 'Understanding face recognition with an interactive activation model', *British Journal of Psychology*, vol.81, pp.361–80.
- Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B. and Henderson, Z. (2001) 'Human and automatic face recognition: a comparison across image formats', *Vision Research*, vol.41, pp.3185–95.
- Diamond, B.J., Valentine, T., Mayes, A.R. and Sandel, M.E. (1994) 'Evidence of covert recognition in a prosopagnosic patient', *Cortex*, vol.28, pp.77–95.
- Diamond, R. and Carey, S. (1986) 'Why faces are and are not special: an effect of expertise', *Journal of Experimental Psychology: General*, vol.115, pp.107–17.
- Ellis, H.D. and Lewis, M.B. (2001) 'Capgras delusion: a window on face recognition', *Trends in Cognitive Sciences*, vol.5, no.4, pp.149–56.
- Ellis, H.D. and Young, A.W. (1990) 'Accounting for delusional misidentifications', *British Journal of Psychiatry*, vol.157, pp.239–48.
- Gibson, J.J. (1986) *The Ecological Approach to Visual Perception*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Hay, D.C. and Young, A.W. (1982) 'The human face', in Ellis, A.W. (ed.) *Normality and Pathology in Cognitive Functions*, London, Academic Press.
- Hay, D.C., Young, A.W. and Ellis, A.W. (1991) 'Routes through the face recognition system', *Quarterly Journal of Experimental Psychology*, vol.43A, pp.761–91.
- Hirstein, W. and Ramachandran, V.S. (1997) 'Capgras syndrome: a novel probe for understanding the neural representation of identity and familiarity of persons', *Proceedings of the Royal Society, London, Series B*, vol.264, pp.437–44.
- Humphreys, G.W. and Bruce, V. (1989) *Visual Cognition: Computational, Experimental and Neuropsychological Perspectives*, Hove, Lawrence Erlbaum Associates Ltd.
- Humphreys, G.W. and Riddoch, M.J. (1984) 'Routes to object constancy: implications from neurological impairments of object constancy', *Quarterly Journal of Experimental Psychology*, vol.36A, pp.385–415.
- Johnson, M.H. and Morton, J. (1991) *Biology and Cognitive Development: The Case of Face Recognition*, Oxford, Blackwell.
- Johnston, A., Hill, H., and Carmen, N. (1992) 'Recognizing faces: effects of lighting direction, inversion and brightness reversal', *Perception*, vol.21, pp.365–75.

- Johnston, R.A. and Bruce, V. (1990) 'Lost properties? Retrieval differences between name codes and semantic codes for familiar people', *Psychological Research*, vol.52, pp.62–7.
- Kemp, R., Pike, G. and Brace, N. (2001) 'Video-based identification procedures: combining best practice and practical requirements when designing identification systems', *Psychology, Public Policy and Law*, vol.7, no.4, pp.802–7.
- Kemp, R., Towell, N. and Pike, G. (1997) 'When seeing should not be believing: photographs, credit cards and fraud', *Applied Cognitive Psychology*, vol.11, no.3, pp.211–22.
- Kilgour, A.R. and Lederman, S.J. (2002) 'Face recognition by hand', *Perception and Psychophysics*, vol.64, pp.339–52.
- Lawson, R. and Humphreys, G.W. (1996) 'View-specificity in object processing: evidence from picture matching', *Journal of Experimental Psychology: Human Perception and Performance*, vol.22, pp.395–416.
- Lederman, S.J. and Klatzky, R.L. (1987) 'Hand movements: a window into haptic object recognition', *Cognitive Psychology*, vol.19, pp.342–8.
- Lederman, S.J. and Klatzky, R.L. (1990) 'Haptic classification of common objects: knowledge-driven exploration', *Cognitive Psychology*, vol.22, pp.421–59.
- Lederman, S.J., Klatzky, R.L. and Pawluk, D.T. (1993) 'Lessons from the study of biological touch for robot haptic sensing', in Nichols, H. (ed.) 'Advanced tactile sensing for robotics', in *World Scientific Series in Robotics and Automated Systems*, vol.5, Singapore, World Scientific Publishing.
- Marr, D. (1977) 'Analysis of occluding contour', *Proceedings of the Royal Society of London*, Series B, vol.197, pp.441–75.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company.
- Marr, D. and Nishihara, H.K. (1978) 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proceedings of the Royal Society of London*, Series B, vol.211, pp.151–80.
- Morrison, D.J., Bruce, V. and Burton, A.M. (2001) 'Understanding provoked overt recognition in prosopagnosia', *Visual Cognition*, vol.8, pp.47–65.
- Neisser, U. (1967) *Cognitive Psychology*, New York, Appleton-Century-Crofts.
- Pike, G., Brace, N. and Kynan, S. (2001) 'The visual identification of suspects: procedures and practice', *A Publication of the Policing and Reducing Crime Unit, Home Office Research, Development and Statistics Directorate*.
- Pike, G., Kemp, R. and Brace, N. (2000) 'The psychology of human face recognition', *IEE Electronics and Communications: Visual Biometrics*, 00/018, pp.12/1–12/6.
- Searcy, J.H. and Bartlett, J.C. (1996) 'Inversion and processing component and spatial-relational information in faces,' *Journal of Experimental Psychology: Human Perception and Performance*, vol.22, pp.904–15.
- Selfridge, O.G. (1959) 'Pandemonium: a paradigm for learning', in *The Mechanisation of Thought Processes*, London, HMSO.

- Sergent, J. and Poncet, M. (1990) 'From covert to overt recognition of faces in a prosopagnosic patient', *Brain*, vol.113, pp.989–1004.
- Tanaka, J.W. (2001) 'The entry point of face recognition: evidence for face expertise', *Journal of Experimental Psychology: General*, vol.130, pp.534–43.
- Tarr, M.J. (1995) 'Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects', *Psychonomic Bulletin and Review*, vol.2, pp.55–82.
- Warrington, E.K. and Taylor, A.M. (1978) 'Two categorical stages of object recognition', *Perception*, vol.7, pp.695–705.
- Yin, R.K. (1969) 'Looking at upside down faces', *Journal of Experimental Psychology*, vol.81, pp.141–5.
- Young, A.W., Hay, D.C. and Ellis, A.W. (1985) 'The faces that launched a thousand slips: everyday difficulties and errors in recognizing people', *British Journal of Psychology*, vol.76, pp.495–523.
- Young, A.W., Newcombe, F., De Haan, E.H.F., Small, M. and Hay, D.C. (1993) 'Face perception after brain injury', *Brain*, vol.116, pp.941–59.

---

# PART 2

## CONCEPTS AND LANGUAGE

---

Introduction

**Chapter 5 Concepts**

*Nick Braisby*

**Chapter 6 Language processing**

*Gareth Gaskell*

**Chapter 7 Language in action**

*Simon Garrod and Anthony J. Sanford*

# Introduction

In Chapter 1, we saw how cognitive psychology seeks to explain cognition in terms of information processing by developing and refining accounts that are expressed in terms of representations, which carry information, and computations, which transform the representations in various ways. Whereas Part 1 showed how such accounts could be developed to explain perceptual processes, in Part 2, we shall see how successfully this approach can be applied to two related areas of cognition: categorization and language.

Categorization, our ability to group things together into discrete categories such as fruit, vegetables, tables and chairs, can be examined in different ways. It can be analysed from a perceptual point of view – how can particular visual or auditory features, for example, influence how we categorize the scenes that we perceive? – but also from a linguistic viewpoint – how is our categorization influenced by the information we receive via language and also by the words we have available? In placing categorization in Part 2, we have chosen to emphasize the link between categorization and language, but in making concepts the topic of the first chapter after Part 1 we also hope to draw attention to some of the links between categorization and perception. Indeed, categorization or semantic classification can be seen as the next stage on from perceptual classification, the focus of Chapter 4.

In Chapter 5, Nick Braisby outlines several different theoretical approaches to categorization. Despite being a fundamental ability, categorization appears to elude a comprehensive treatment. The first two theories outlined, classical and prototype theories, imply that concepts, our mental representations of categories, can be neatly demarcated one from another, and each understood in terms of sets of features. According to both theories we place items into a category if they possess a criterial number of these features. Such theories are knowledge-lean – that is, they assume first that it is possible to demarcate category-relevant knowledge, and second that only this knowledge is relevant to determining categorization.

However, both of these theories suffer a number of problems. The alternative theories discussed in the chapter assume that categorization is knowledge-rich, that is, it involves broader knowledge structures – lay theories about domains are implicated by the ‘theory’-theory of concepts, and beliefs about what constitute essential properties are implied by psychological essentialism.

As broader knowledge structures get invoked to explain categorisation, however, you will see that it becomes harder to state theories precisely, and the discussion of theories in the chapter reflects this. Whereas classical and prototype theories are outlined with some precision, so that one can imagine detailed accounts of the process of categorization being given, ‘theory’-based and essentialist theories are hard to define, and it is unclear whether an information processing account could be developed at present.

Because of the difficulty in developing precise accounts of representations of categories and the processes constituting categorization, researchers have been led to revisit some of the simplifying assumptions previously made in this literature. Perhaps, for example, there might be different kinds of categorization for different kinds of category, or for different kinds of categorizer. In some sense, researchers are

considering again what categorization really is. In the terminology of Marr's levels that we saw in Chapter 1, in spite of its fundamental importance, researchers are still seeking agreement over a level 1 account of categorization. Only then might we hope to develop precise level 2 accounts.

Gareth Gaskell's Chapter 6 builds on some of the foundations of Chapter 5, but seeks to explain something that superficially appears very different – how we comprehend both spoken and written language. Together with Simon Garrod and Anthony Sanford's Chapter 7, these chapters on language mark a point of departure. The chapters in Part 1, and Chapter 5 to some extent, have been concerned with how we perceive and pick up information concerning our environment, and how we use this information to infer the presence and nature of objects, and the categories to which they belong. Chapters 6 and 7 mark a concern with the social world, with how we communicate about our world to others, how we make sense of the interpretations of others, and how others influence our communication. As Gareth Gaskell states in opening his chapter, understanding our ability to use language is key to understanding what differentiates humans from other animals, and key to understanding human cognition.

Chapter 6 draws our attention to many aspects of language processing that we normally take for granted. In comprehending spoken language, we have to infer which words are present in a stream of speech, an ability we learn as children. We also have to learn to make use of our knowledge of the speech sounds used in our particular language(s). These processes are easily taken for granted, and researchers have had to coin new terms and posit new theoretical structures, such as the mental lexicon, in order to make sense of the comprehension process. Researchers have assumed that different kinds of knowledge are brought to bear at different stages of comprehension. Models of the process that incorporate new theoretical structures and different kinds of knowledge have been constructed (e.g. the cohort model) and experiments conducted to evaluate them. Indeed, and in contrast with Chapter 5, Chapter 6 focuses mainly on processing accounts and how well they explain experimental data. Also in contrast with Chapter 5, some of the processing accounts are sufficiently well specified that they have been developed as computational models. TRACE, for example, is a connectionist model, as is IAC, a model similar to one you saw in Chapter 4. That researchers have been able to develop such models successfully is a testament to how advanced is our understanding of the cognitive processes of language comprehension.

Nevertheless, running through Chapter 6, and in common with Chapter 5, is a concern with the extent to which we require general knowledge for processing language, and the time-point at which this knowledge is brought to bear. The bottom-up (and autonomous) view is that, for example, semantic knowledge is only called upon late on in processing, and only then to adjudicate between interpretations of the incoming input. The top-down (and interactive) view is that such knowledge may operate early, and influence which interpretations are pursued. This important debate, ranging over phenomena such as spoken and written word recognition, ambiguity resolution, and sentence comprehension, is as important as it is unresolved.

This debate is also reflected in Chapter 7, where Simon Garrod and Anthony Sanford broaden the focus on language to include the comprehension of whole texts

(not just sentences), language production, and dialogue. The authors begin by considering some difficulties for a simple view of language processing in which information is processed at one discrete stage, and then passed to another, and to another. On the simple view, each stage involves interpreting or translating the output of the previous stage. So, for example, in speech comprehension, a phonological stage might be followed by a lexical stage, which might be followed by a syntactic stage, and so on. Each stage takes as its input the output of the previous stage, transforms that input in a certain way, and then outputs to a subsequent stage. The theme that runs throughout Chapter 7 is that understanding and producing language involves much more than the simple view assumes – in particular, the authors show how language processing as a whole involves drawing heavily on general, non-linguistic knowledge. This is a significant contrast with Chapter 6, where it was assumed that knowledge could be neatly compartmentalized, and that language comprehension required the use of only particular kinds of knowledge and only at particular stages.

Simon Garrod and Anthony Sanford begin by showing how the comprehension of texts does not just rely on consulting the meaning of words in the mental lexicon, and combining these according to linguistic rules. In various ways, texts require the application of much more than just lexical knowledge – for example, the authors suggest that the meaning of some words is unlikely to be represented in a ‘lexicon’ but rather is rooted in actual bodily posture and movement. That language can involve more complicated processes is also demonstrated by a discussion of language production. Though there are some similarities to the discussions in Chapter 5 – production can be seen as involving the reverse of comprehension processes – there are also further complexities. Production involves monitoring one’s productions to ensure they are as intended, and also designing one’s productions according to the social context. And it is the social aspect of language that most clearly comes to the fore in the discussion of dialogue. Dialogue involves coordination between speakers at a number of levels: for example, in taking turns to ask and answer questions, in developing a common understanding, and in what the authors describe as alignment and routinization of representations.

What do these three chapters reveal about the cognitive approach? Perhaps most notable in these chapters is the breadth of the cognitive approach. Researchers tackle very diverse questions – from what knowledge we have of categories to what processes underpin dialogue – but do so from a common perspective – that of seeking to posit mental representations, which carry information, and computational processes that transform them.

The chapters also invite us to think about the success of the cognitive approach. Chapter 6 shows how cognitive psychology has been successful in generating detailed processing models of language comprehension. Chapters 5 and 7 show a different kind of success – though researchers have yet to answer some of the basic questions about categorization and language-in-action, the cognitive approach has helped them to generate different theoretical frameworks and empirical means of examining them. That is, the success of the approach can be measured not only in terms of the success of proposed models, but also in terms of the generation of new research questions.

Finally, the three chapters in this part reveal a reciprocity between the precision with which theories and models may be specified and the extent to which a cognitive process appears to be knowledge-rich, the extent to which it seems to draw on general knowledge. The more general knowledge a process draws upon, the harder it is for researchers to develop precise models. It appears that precision – one of the hallmarks of a scientific account – can be achieved only when the knowledge that influences a cognitive process can be isolated or separated from other kinds of knowledge and demarcated in distinct processing modules. The question as to the modularity of cognition, explicitly addressed in Chapter 17 by Tony Stone, is one to which we shall return again and again.



Nick Braisby

---

### 1 Introduction

In the UK some years ago a television channel screened a programme that involved contestants guessing the identity of unusual antique artefacts. The contestants were allowed to hold the objects, and discuss their ideas as to what they might be. But the objects were chosen so that it was not at all obvious what they were used for, nor what they were called. In the parlance of cognitive psychology, soon to be explained, they were selected because they were difficult to categorize – they were objects for which the contestants could not readily find an appropriate concept. You can get an idea of the difficulty faced by these contestants from Activity 5.1.

#### ACTIVITY 5.1

---

Figure 5.1 shows some obscure artefacts that may be found in the kitchen. Try to guess what these objects are – answers are given at the end of the chapter.



**Figure 5.1** Three (more or less) obscure objects that may be found in the kitchen

---

Normally we categorize things effortlessly. Looking around me now I can't see a single object that I can't label or categorize – I don't have to think hard to identify the appropriate concept for each and every one. But how do we do this? For, as Activity 5.1 shows, as did the television programme, categorizing something, finding the right concept, can be difficult. In fact, as we shall see, even effortless categorizations are ultimately difficult to explain. The first step is to work out what concepts are.

#### 1.1 Concepts, categories and words

Dictionaries say that the word ‘concept’ has different senses. There is a non-technical sense, one that relates loosely to ideas and thoughts. So, for example, we might say that a manager has created a new marketing ‘concept’, meaning he or she has introduced a new idea for promoting a product. However, it is the psychological or philosophical sense that is of interest here. According to this, **concepts** are general ideas formed in the mind: ‘general’ meaning that concepts apply to every one of a class of things (usually described as a category). For example, my concept of ‘cat’ must be a general idea of cats – an idea of what cats are in general, that is, an idea about all cats, not just my pet cat Rosie curled up on the sofa.

Already this raises two important issues. First, concepts are related to **categories**. Our talk of concepts normally presupposes the existence of a corresponding category. There are similarities here with the discussion of Brentano in Chapter 1. Brentano argued that a mental state has two components – a mental act, internal to the mind, and a mental content (the thing that the mental act is about) that is *external* to the mind. Concepts also have this dual aspect. In thinking ‘Rosie is a cat’ I perform an activity (thinking) and my thought has a content that is external to the mind – the thought is about Rosie and her relation to the category of (domestic) cats. So, although concepts are internal to the mind, the categories that concepts are about are external. Indeed, researchers often adopt the terminological distinction that the word ‘concept’ refers to something in the mind and ‘category’ refers to those things in the world which a concept is about.

Second, concepts and categories are linked to words. I used words to communicate the idea that thoughts (such as Rosie is a cat) contain concepts, and that concepts are about categories. Words like ‘cat’ help you to work out what concept I have in mind (the concept ‘cat’). However, it would be too simplistic to suggest that words always pick out concepts straightforwardly. Ambiguous words link to more than one concept – ‘chest’ relates both to the concept of a body part (torso) and to the concept of furniture (as in chest of drawers). In addition, most words are polysemous – they have many distinct but closely related senses. ‘Cat’ can refer to the category of domestic cats, but also to big cats and to all felines. Concepts, unlike words, do not have multiple senses, since they are general ideas about particular categories. So, we probably have several concepts that all link to the word ‘cat’ – a concept of ‘domestic cat’, a concept of ‘feline’, and so on. Mapping the precise relationships between concepts and words is not at all easy, so for much of this chapter I will assume, as most researchers do for practical reasons, that words pick out concepts in a straightforward manner. Towards the end of the chapter, though, I will try to show some of the complexity of this relationship.

Having considered some preliminaries, we can now turn to the kind of evidence psychologists have used to infer the nature of concepts.

## 1.2 Categorization

Bruner *et al.* (1956, p.1) suggested that ‘to categorize is to render discriminably different things equivalent, to group objects and events and people around us into classes, and to respond to them in terms of their class membership rather than their uniqueness’. According to this definition, concepts are at work whenever people show similarities in behaviour toward different objects and whenever they show differences in behaviour toward different objects. If, for example, you pat two different dogs, you behave similarly towards them, in spite of their differences. On these definitions you do so because you treat them as instances of the category ‘dog’. Likewise, patting a dog but not a house plant signals that you treat these as members of different categories.

Even though concepts may be at work almost all of the time, our focus will be on a restricted range of behaviours that involve giving fairly explicit and often linguistic judgements about category membership. This kind of categorization behaviour (henceforth, ‘categorization’) has provided the primary evidence as to the nature of concepts.

Categorization behaviour could be more broadly construed however. Potter and Wetherell (1987), and Edwards and Potter (1992) show how attention to natural discourse reveals many subtleties in how people choose which category words to use, and how they then use them in particular contexts. This ‘discursive’ approach can show how categorization is affected by social influences, such as the social status of the people discoursing, and how using category words serves broader goals than merely that of reporting one’s beliefs about category membership. Though this line of work reveals important aspects of categorization, cognitive psychologists are interested in what we can learn about categorization processes in general; that is, in what might be common to different instances of categorization in different contexts.

Categorization behaviour also need not be so closely tied to language. Indeed, many researchers attribute concepts to non-linguistic animals. Sappington and Goldman (1994) investigated the abilities of Arabian horses to learn to discriminate patterns. They claimed that horses that learned to discriminate triangles from other shapes had actually acquired a concept – in this case, the concept of ‘triangularity’ – as opposed to merely having learned the particular triangular patterns to which they had been exposed. Again, though, this chapter focuses on what we know of human cognitive achievements, and hence on the nature of human concepts.

My bracketing-off of these two issues does not solely reflect a pragmatic desire to get on with discussing the matters of most relevance to cognitive psychology, but also the contentious nature of these issues. For instance, some cognitive scientists have argued that the idea that animals might possess concepts is not actually coherent (Chater and Heyes, 1994). Similarly, others have argued against a strong discursive position, according to which categories are socially constructed (e.g. Pinker, 1997; Fodor, 1998).

So, accepting that judgements of category membership are the principal indices of categorization, we can now turn to some of the techniques psychologists have used to elicit these. One method is the **sorting** task. In this task, participants are shown an array of different items (sometimes words printed on cards) and asked to sort them into groups. Ross and Murphy (1999) used this technique to examine how people categorize foods. They found, for example, that people sometimes put eggs in the same group as bacon and cereal (suggesting a category of breakfast foods), whereas at other times they put eggs together with butter and milk (suggesting a category of dairy products). The groups into which items are categorized are taken to reflect corresponding concepts. The fact that eggs are sometimes put into different groups is consistent with Barsalou’s (1983) findings (and also with the discursive view) that categorization depends upon people’s goals or purposes. So, for example, when asked to say what falls into the category ‘things to take with you in case of fire’ people would mention items that would not normally be found together in the same category (e.g. loved ones, pets and family heirlooms).

If the sorting task seems abstract and artificial, go into your kitchen and look at how the different foods and gadgets are organized. You will probably find items grouped into categories – herbs and spices in one place, for example, fruit in another, vegetables in yet another. I group foods together even in my supermarket carrier bags – usually into nothing more complicated than frozen, chilled and room temperature – when the person at the checkout gives me enough time to do so! So, placing

members of a category together is really an everyday activity that the sorting task taps into in a measurable and controlled way.

## ACTIVITY 5.2

---

Think of the properties that dogs have. You might think dogs ‘bark’, ‘have four legs’, ‘run after sticks’, ‘pant’, and so on. Now consider the concept of ‘cat’. Take two minutes to write down some of the properties cats have. Don’t dwell on any particular property: just write down whatever comes to mind. If you get to 10 properties, stop.

### COMMENT

Simple though this task seems, it gets hard to think of new properties after a while. Psychologists use this **property-listing** (or **attribute-listing**) technique to investigate people’s concepts, obtaining results from many participants for each category. They then compare the lists from different people and generate a further list of the most frequently mentioned properties. This gives an indication of the information incorporated in people’s concepts, and the frequency of mention indicates how central each property is to the concept.

---

The sorting task and property-listing technique highlight a third aspect of concepts – they are invoked to explain categorization behaviour. We behave differently with cats and dogs, because cats and dogs belong to different categories, and so our concepts of cats and dogs must differ. The differences (and similarities) between cats and dogs are reflected in our concepts, and it is these concepts that are involved in producing our behaviour.

## 1.3 The wider story of concepts

Perhaps because concepts are implicated in so much of our behaviour, their role often goes unnoticed. However, there have been times when concepts have been the explicit focus of discussion. Umberto Eco (1999) discusses the example of the platypus. In 1798 a stuffed platypus was sent to the British Museum. Initially, it was considered so strange that it was thought to be a hoax, with its beak artificially grafted onto its body. For the next eighty years the question of how the platypus should be categorized was hotly debated. Finally, in 1884, it was declared to be a type of mammal, called a ‘monotreme’, which both lays eggs and suckles its young, and this categorization has stuck (though, of course, as you will see in Section 2.1.4, it is conceivable that even this categorization might again come into question).

This case of scientific ‘discovery’ reminds us that all of our concepts have a past. Even such basic concepts as ‘human’, ‘table’ and ‘food’ have a rich, though perhaps not fully discoverable, history. But the example of the platypus shows that categorization can be a very complex process. Even though everyday categorizations seem effortless and routine, it took the best scientific minds nearly 90 years to decide how the platypus should be categorized.

Box 5.1 offers another example of where categorization has been more explicitly discussed; legal and moral cases provide others. In the UK, for example, the law applies differently to adults and children. So, it is important to be able to categorize

everyone as either a child or an adult. Yet, it is too difficult to identify a precise age for the boundary between children and adults, and so parliaments have to decide, arbitrarily, where it should lie.

## 5.1

### Categorization and diagnosis

Clinicians need to categorize conditions and diseases in order to treat their patients. Though we usually call this diagnosis, it is really a form of categorization – clinicians consider the various properties or symptoms that a patient manifests, and attempt to categorize or diagnose the underlying condition. For example, diagnosing or categorizing chronic fatigue, or ME, is notoriously difficult. Macintyre (1998) suggests diagnosis should be based on major criteria – chronic unexplained fatigue that is debilitating, and which is not due to exertion, nor substantially alleviated by rest. She also suggests that at least four out of eight minor criteria should be present (e.g. sore throat, muscle pain).

Categorization can also be seen in the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994), which gives criteria for diagnosing different mental illnesses. For example, a diagnosis of schizophrenia should be made on the basis of characteristic symptoms, social or occupational dysfunction, duration, and so on. Although the manual lists five characteristic symptoms (e.g. delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behaviour), it indicates that a diagnosis of schizophrenia may be made when only two are present.

You will see later that both of these kinds of diagnosis, which require only a certain number of a longer list of symptoms to be present, relate to a particular theoretical approach to concepts. Though our discussion of concepts is rooted in laboratory-based studies, it is just a short step to matters of practical import.

Fascinating though these examples are, the rest of the chapter concentrates on more everyday categorizations. Researchers have tended to adopt a methodological strategy of explaining the simpler cases first, in the hope that explanations can then be developed for more complex cases. As you will see, even everyday categorizations are surprisingly difficult to explain.

## 1.4 Concepts and cognition

In the last chapter, you saw that the word ‘recognition’ labels different kinds of process. The authors focused on what was called ‘perceptual classification’ and you may have wondered about the subsequent stage labelled ‘semantic classification’. Well, semantic classification is what concepts are all about. So, the use of concepts to classify – for example, using the concept of ‘cat’ to classify or categorize my pet cat Rosie – can be viewed as a further kind of recognition.

Concepts can also be seen as the basic units of semantic memory. While episodic memory stores memories of particular episodes, such as what happened on your last birthday, semantic memory is our long-term memory for facts about the world such as ‘cats are animals’. The episodic–semantic distinction, which you have already

met in Chapter 2, is discussed in more detail in Chapter 8. For our purposes, we simply note that elements of semantic memory such as ‘cats are animals’ express relationships between concepts (between ‘cat’ and ‘animal’ in this case).

We have already mentioned the relationship between concepts and words, but many researchers assume a more explicit link. It is thought that some concepts, called **lexical concepts** (i.e. concepts for which there is a single word), represent our understandings of the meanings of words and are stored in something called the mental lexicon (see Chapter 6). For example, our concept of ‘cat’ would represent what we believe the word ‘cat’ means. The process of understanding language therefore partly involves retrieving lexical concepts from the mental lexicon. Of course, this is a complex process: there may be several lexical concepts corresponding to a single word like ‘cat,’ so we would also have to identify which lexical concept is most appropriate. These and other complexities are developed in Chapter 6.

Concepts also play a role in reasoning. Your list from Activity 5.2 indicates some of the information in your concept of ‘cat’. You may have written things like ‘meows’, ‘likes fish’, ‘mammal’, and so on. You might not have written ‘has a heart’ but this is a property of cats too. Now suppose someone asked you whether Rosie has a heart. My guess is that you would say she does. But this is curious, because I have told you only that Rosie is a cat. How have you managed to draw the inference that she has a heart? The answer, of course, is that your concept of ‘cat’ indicates that cats are mammals, and your concept of ‘mammal’ indicates that mammals have hearts. From these concepts you can infer that cats, like Rosie, have hearts. Such inferences might not always be valid of course – though I don’t doubt Rosie has a heart, for all I know, maybe, miraculously, she has some complex artificial pump instead. The complexities of reasoning, of drawing inferences, are the topic of Chapter 10.

Because concepts allow us to make inferences, they simplify the task of remembering information. If you want to remember the properties of Rosie, you would do well simply to remember that she is a cat. If she were unusual (such as having a piece of her ear missing), you might have to remember that information too. But you do not need to remember explicitly that Rosie meows, or that she has a heart, because you can draw these inferences simply by knowing she is a cat. Suffice it to say that our ability to store concepts in semantic memory, together with our ability to reason and draw inferences, simplifies the task of remembering information. Here, concepts, reasoning and memory all act together.

## Summary of Section 1

---

- Concepts are ideas in the mind that are about categories in the world.
- Words tend to pick out concepts, though the exact relationship between them is complex.
- The principal evidence for concepts comes from categorization behaviour, which involves people making judgements concerning category membership.
- Concepts play a wide role in cognition, being involved in recognition, language, reasoning and semantic memory, to name but a few.

## 2 Explaining categorization

How do we decide that some items belong to the same category and other items belong to different categories? What is it about different cats, for example, that makes us think they are all ‘cats’ and not ‘dogs’?

### 2.1 Similarity I: the classical view of concepts

According to the **classical** view of concepts, which has its roots in the philosophical writings of Aristotle (Sutcliffe, 1993), things belong to categories because they possess certain properties in common. There are two aspects to this idea. First, if something is a member of a category, then it must possess the properties common to the category’s members. Second, if something possesses the properties common to a category’s members, then it too must be a member of the category. The first aspect asserts that possession of the common properties is necessary for category membership; the second indicates that possession of the common properties is sufficient for category membership. The classical view, then, is that there are both *necessary* and *sufficient* conditions on category membership. Another way of expressing this is to say that the classical view is that concepts provide **definitions** of their corresponding category.

In this view, categorization is explained in terms of a comparison of any putative instance with the conditions that define the category. If the instance matches the concept on each and every condition, then it falls within the category – it is a member of the category. If it fails to match on any condition, then the instance falls outside the category – it is a non-member. Let’s consider an example – the category of bachelors. The classical view contends that the category can be defined, that there are properties that are both necessary and sufficient for membership. Might this be true? Dictionaries tell us that bachelors are unmarried, adult males. Perhaps these properties are necessary and sufficient for bachelorhood. If they are, then any person who is a bachelor must also be unmarried, adult and male. Conversely, any person who is unmarried, adult and male must be a bachelor. And this seems right: it doesn’t seem possible to imagine a bachelor who is married, say. Nor does it seem possible to imagine someone who is unmarried, adult and male who isn’t a bachelor.

---

#### ACTIVITY 5.3

Consider the categories sparrow, gold, chair, introvert, red, and even number. Can you provide definitions for them? Take a few minutes to list the properties for each that you think are important for category membership. Don’t worry if you find this difficult: just write down what comes to mind. If you can’t think of anything, pass on to the next category. When you have finished try to answer the following questions. First, do you think each of these properties is necessary for category membership (i.e. must every member of the category possess the property)? Second, are the properties for each category, when taken together, sufficient for membership in the category (i.e. must anything that possesses these properties necessarily be a member of the category)?

## COMMENT

Most people find this kind of activity difficult. In spite of the classical view, it is surprisingly difficult to think of watertight definitions – you might have succeeded for ‘even number’ but perhaps not for the other categories. We will consider this again in Section 2.1.4.

---

The classical view was supported by some early, empirical investigations (e.g. Hull, 1920; Bruner *et al.*, 1956) that showed people categorized instances according to whether they possessed the necessary and sufficient conditions of the category. However, despite being sporadically defended (e.g. Sutcliffe, 1993), there have been numerous criticisms. The first concerns the phenomenon known as ‘typicality’.

### 2.1.1 Typicality

Since the classical view contends that all members of a category must satisfy the same definition, it follows that they should all be equally good members of that category. However, psychologists have found systematic inequalities between category members. Rosch (1973) elicited participants’ ratings of the typicality or ‘goodness-of-exemplar’ (sometimes referred to as GOE) of particular instances of a category – the method is often known as a **typicality ratings** method. Rosch’s instructions give a sense of what is involved.

Think of dogs. You all have some notion of what a ‘real dog,’ a ‘doggy dog’ is. To me a retriever or a German shepherd is a very doggy dog while a Pekinese is a less doggy dog. Notice that this kind of judgement has nothing to do with how well you like the thing ... You may prefer to own a Pekinese without thinking that it is the breed that best represents what people mean by dogginess.

(Rosch, 1973, pp.131–2)

---

## ACTIVITY 5.4

Now that you have read Rosch’s instructions, write down the following words on the left-hand side of a sheet of paper, putting each word on a new line: pineapple, olive, apple, fig, plum, and strawberry. Then, to the right of each word, write down the number (between 1 and 7) that best reflects how well the word fits your idea or image of the category ‘fruit’. A ‘1’ means the object is a very good example of your idea of what the category is, a ‘7’ means the object fits very poorly with your idea or image of the category (or is not a member at all).

## COMMENT

When you have written down your answers, compare your ratings with those of Rosch shown in the first column of Table 5.1 (see top of page 172). How might you explain these ratings? Many people feel that their ratings reflect how familiar they are with particular instances, or how frequently those instances are encountered.

You might think that in a society where figs were more commonplace than apples, for example, the typicality of these items would be reversed. In a series of studies, Barsalou (1985) investigated the influences of familiarity and frequency on typicality. Contrary to what one might think, he found that typicality did not correlate with familiarity, and only correlated with frequency to a limited extent. So, it seems that even if penguins were much more common in our lives than they are, and we were all much more familiar with them, we would still think of them as atypical birds!

---

Rosch's results for four different categories are shown in Table 5.1 (overleaf). She took these ratings to be indicative of the internal structure of categories, and this conclusion was supported by other empirical work. For instance, Rips *et al.* (1973) and Rosch (1975) examined the relationship between typicality and the time it takes participants to verify sentences that express categorization judgements. The method is often known as **category** or **sentence verification**. For example, the sentences might be 'a robin is a bird' (typical instance) and 'a penguin is a bird' (atypical instance). Participants were asked to respond either 'Yes' (meaning they thought the sentence was true) or 'No' (meaning they thought it was false) as quickly as possible. The results showed that for highly typical sentences people were much quicker to verify the sentence (i.e. the sentence 'a robin is a bird' was verified more quickly than the sentence 'a penguin is a bird').

Further support for the idea that categories have internal structure came from Rosch and Mervis (1975). They used the property- or attribute-listing method, the method you tried in Activity 5.2. They asked their participants to generate lists of properties for a series of category instances, for example, robin and penguin for the category bird. The results showed that less typical instances shared properties with fewer category members, while more typical instances shared properties with many other instances. For example, robins have properties – flying, eating worms, building nests – that are shared with many other birds. Penguins have properties – swimming, not flying – that are shared with relatively few other birds.

Using methods such as these, Rosch, Mervis and others provided impressive evidence that categories have what we might think of as a rich **internal structure**. A definition serves to demarcate members of a category from non-members, but even things inside the category are highly structured. Both penguins and robins would satisfy the definition of a bird, but there are important systematic differences between them that are reflected in the cognitive processes governing categorization. And this seems contrary to the classical view's suggestion that all category members must equally satisfy a category's definition. How can categories have highly typical and atypical members if the classical view is correct? And how strongly does such evidence count against the classical view?

Though the classical view makes strong claims about the membership of categories – membership should be all-or-none – it says nothing about their internal structure. So, the findings of rich internal structure do not show the classical view to be wrong, unless, of course, internal structure reflects category membership. If a penguin were not only a less typical bird than a robin, but also less of a

**Table 5.1** Rosch's (1973) typicality ratings for various instances of four categories

Fruit		Vegetable		Sport		Vehicle	
Apple	1.3	Carrot	1.1	Football	1.2	Car	1.0
Plum	2.3	Asparagus	1.3	Hockey	1.8	Scooter	2.5
Pineapple	2.3	Celery	1.7	Gymnastics	2.6	Boat	2.7
Strawberry	2.3	Onion	2.7	Wrestling	3.0	Tricycle	3.5
Fig	4.7	Parsley	3.8	Archery	3.9	Skis	5.7
Olive	6.2	Pickle	4.4	Weight-lifting	4.7	Horse	5.9

category member than a robin, then ratings of typicality might reflect a graded notion of category membership in which categories have some clear members, some clear non-members, and a range of cases in between. Then, category membership, quite palpably, would not be all-or-none. On the other hand, if typicality does not reflect graded membership, it may be compatible with the classical view. However, typicality effects do expose an inadequacy in the classical view, even if they do not contradict its basic tenets. It is not at all obvious how the classical view might explain typicality effects; at the very least, it would need supplementing.

### 2.1.2 Borderline cases

If membership in a category is ‘all-or-none’, as the classical view suggests, then there should be no borderline cases: an item either satisfies the definition of a category or it doesn’t. Intuition alone tells us there are items whose category membership is unclear. Colour categories, for example, have no obvious boundary. It seems impossible to draw a line on the colour spectrum, say, between red and orange. For where does a red shade fade into orange? Rather, in between these two categories, there seem to be shades that are neither unequivocally red nor unequivocally orange, hence our use of phrases such as ‘a reddy-orange’.

McCloskey and Glucksberg (1978) provided evidence that confirmed this intuition for a whole range of categories. They used a method of asking for **categorization judgements**. They asked their participants to respond either ‘Yes’ or ‘No’ to questions of category membership (such as ‘Is a robin a bird?’). Participants were also asked to rate the same instances for typicality. McCloskey and Glucksberg then considered the level of agreement that participants showed in their categorization judgements, both across individuals and within the same individuals over two times of testing. They found that participants readily agreed on highly typical and atypical items, yet disagreed over time and across individuals for some items of intermediate typicality. For example, people rated ‘chair’ as a highly typical item of ‘furniture’, and were consistent amongst themselves, and over time, in judging a chair to be an item of furniture. Similarly, with highly atypical items such as a ceiling, they were consistent in judging this not to be an example of furniture. With items of intermediate typicality, such as bookends, they were much less

consistent. Some people judged these to be items of furniture, others did not; and some people changed their judgements across the two times of testing. McCloskey and Glucksberg thus gave empirical weight to the intuition that many categories have borderline cases.

How telling is this evidence? The classical view certainly implies that categories should have no borderline cases. However, it is at least possible that some of the instances, which appear borderline, are not genuinely indeterminate, unlike the case of colour categories. It might be that patterns of disagreement reveal a lack of knowledge. For example, you may not know whether a tomato is a fruit or a vegetable. Perhaps sometimes you will say it is a fruit, other times you might say it is a vegetable. But, if you consult a dictionary, you will be told that it is a fruit, even though it is usually used as a vegetable (e.g. in sauces). So, it is possible that an instance definitely belongs to one or other category (i.e. is not borderline), but uncertainty makes the item appear borderline. Another possibility is that inconsistency reflects perspective-dependence. It might be, for example, that you know that a tomato is technically a fruit, but your categorization judgement is influenced by the fact that it is used mostly as a vegetable. So, you might agree, in a culinary context, that a tomato is a vegetable, but disagree in the context of a biology lesson.

Though these remain logical possibilities, it is not obvious that McCloskey and Glucksberg's examples actually did involve uncertainty or perspective-dependence. Though people disagreed about whether bookends count as furniture, it seems implausible that they did not have enough information or were adopting different perspectives. So, in the absence of alternative explanations, the compelling evidence for borderline cases seems to undermine the classical view.

### 2.1.3 Intransitivity of categorization

A further source of difficulties for the classical view has been the observation of intransitivity in categorization judgements. Transitivity is observed with many relationships: the relation 'taller than' is transitive because if 'A is taller than B' and 'B is taller than C', then it simply follows that 'A is taller than C'. The relationship is 'transitive' because the last statement follows from the first two.

Is categorization transitive? That is, if As are members of category B, and Bs are members of category C, does it follow that As are also members of category C? According to the classical view it does (and perhaps your intuition agrees). As you have seen, the classical view holds that membership in a category is all or none – if an instance falls into a category, it does so unequivocally. So, if rabbits are mammals then, according to the classical view, they possess the defining features of mammals, and so are mammals unequivocally. Likewise, if mammals are animals, then they possess the defining features of animals, and so are animals unequivocally. There can be no exceptions. So it should just follow, unequivocally, that rabbits must also be animals.

Hampton (1982), however, showed that people's categorization judgements are not in general consistent with transitivity. For example, he found that participants would agree that 'car seats are a kind of chair' and that 'chairs are a kind of furniture' but not agree that 'car seats are a kind of furniture'. Similarly, people might agree that

Big Ben is a clock and that clocks are furniture, but not that Big Ben is an item of furniture. The fact that people strongly reject the transitive inference in these cases represents a real problem for the classical view.

#### 2.1.4 The lack of definitions

In developing his account of language-games, Wittgenstein (1953) considered the idea, as implied by the classical view, that there are common properties to all instances of the category of game:

Consider for example the proceedings that we call ‘games’. I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? – Don’t say: ‘There *must* be something common, or they would not be called “games”’ – but *look and see* whether there is anything common to all. – For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. ...

I can think of no better expression to characterize these similarities than ‘family resemblances’; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. – And I shall say: ‘games’ form a family.

(Wittgenstein, 1953, paras 66–7)

If Wittgenstein is right, then the classical view is simply mistaken. Whereas it contends that categories have common properties, Wittgenstein’s position is that most categories are like ‘game’ – when you look closely for common properties, you find none. Recall Activity 5.3: there you tried to offer definitions of categories such as red, and introvert. Most people find this task difficult, except perhaps for ‘even number’, where there is a rule that defines category membership. Wittgenstein suggests that most categories are really indefinable. Indeed, his position makes sense of a striking anomaly: despite the classical view having a long history, people have identified very few examples of categories that can be defined. Most researchers are forced to fall back on one of a very few examples – my choice of ‘bachelor’ is particularly hackneyed! I couldn’t use another example, such as ‘tree’, ‘river’, ‘chair’ or ‘ship’ because no-one has identified watertight definitions for these categories.

Nonetheless, Wittgenstein has not proved that natural categories cannot be defined, and so it is possible that someone might yet provide definitions. But the philosophers Kripke (1972) and Putnam (1975) undermined even that idea. They considered what would happen if something that was taken to be ‘definitional’ was later found to be wrong. Consider the concept of ‘cat’. Most people would say that cats are mammals, that they have fur and meow, and so on. Are these necessary properties of the category? Well, perhaps there are some cats that don’t meow, some that don’t have fur, but surely all cats are mammals – almost by definition one might say. Putnam considered the implications of discovering that all cats

are really robots controlled from Mars (i.e. not mammals at all). This is a thought experiment, of course, so don't worry that the scenario is improbable, or even impossible. The critical issue is what would be the implications of such a discovery. In particular, would the things that we had previously called cats still be cats? What do you think? If you had a pet cat, would it still be a 'cat' after this discovery? Kripke and Putnam believe that it would – those things we called cats before the discovery are still cats afterwards (i.e. the robots controlled from Mars are still cats). But since robots aren't mammals, 'being a mammal' could not be a defining feature of cats, even though we previously thought it was! The conclusion that Kripke and Putnam draw is that we might be shown to be wrong about virtually any property that we happen to believe is true (or even defining) of a category.

This is how Pinker puts it:

What is the definition of *lion*? You might say 'a large, ferocious cat that lives in Africa.' ... Suppose scientists discovered that lions weren't innately ferocious ... Suppose it turned out that they were not even cats ... you would probably feel that these ... were still really lions, even if not a word of the definition survived. Lions just don't *have* definitions.

(Pinker, 1997, p.323, original emphasis)

There are less fanciful examples that convey the same point. As you saw in Section 1.3, people thought the platypus bizarrely combined the features of birds (a bill), amphibians (swimming), and mammals (fur). Suppose that some people came to believe, erroneously, that the platypus really was a strange kind of bird. What Kripke and Putnam argue is that in a case like this, no matter how strongly held the belief, it could never be definitional for these people that a platypus is a bird. If it were, then as soon as it was determined that the platypus was a mammal after all, by the very same definition it would no longer be a platypus. The arguments of Kripke and Putnam hinge on the intuition that the platypus will still be a 'platypus' no matter what we come to believe, and no matter how wrong those beliefs ultimately turn out. If so, then our beliefs about natural categories never really amount to definitions and the classical view must be mistaken.

## 2.2 Similarity II: prototype theories of concepts

The combined weight of evidence calling into question the classical view led researchers to consider alternatives. Observations of typicality effects suggested to some that concepts are organized around a measure of the central tendency of a category, known as the **prototype**. Sometimes the prototype may correspond to an actual instance, but in general it is like a 'best' category member, formed by statistically aggregating over examples of the category one encounters. Rosch, for instance, believed that it is a feature of the natural world that certain attributes or properties tend to correlate or cluster together, and it is these natural clusters of correlated attributes that prototypes describe. For example, the prototype for 'bird' might describe the cluster of properties such as having feathers, wings, a beak and an ability to fly. These properties cluster together in a way that feathers, lips, gills, and an ability to swing through tree branches do not. Whether or not an instance is a

category member then depends upon how similar it is to the prototype: an instance falls within the category if it achieves a certain criterion of similarity. If an instance is too dissimilar, it mismatches on too many properties, then it falls outside the category.

This account is a little like the classical view: both are committed to the idea that similarity explains categorization. For classical theory, instances fall within a category if they match each and every element of the category's definition, and outside the category if they mismatch on any one. The critical difference is that for prototype theories an instance may fall within a category even if it mismatches on a number of properties. Though it might not seem dramatic, a simple illustration shows how significant a move this really is. Suppose a category is characterized by five properties (call them A, B, C, D and E). Now suppose that there is a criterion for membership in the category such that an instance can mismatch on up to, but no more than, two of these five properties. Then there are a number of logical possibilities for category membership, as shown in Table 5.2.

**Table 5.2** Different kinds of instances (1 to 4) for a category with five characteristic properties. A tick implies an instance matches on a particular property; a cross implies a mismatch

Instances	Properties				
	A	B	C	D	E
1	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✗
3	✓	✗	✓	✗	✓
4	✗	✓	✗	✓	✓

Instance 1 possesses all the characteristic properties of the category. No instance could match on more properties, and so we could think of this as a highly typical, perhaps even a prototypical, instance. Instance 2 mismatches on one property and so is less typical. Instances 3 and 4 mismatch on two properties and are less typical again. What Table 5.2 shows is that this category could not be given a simple definition in terms of the five properties: for each property A to E there is an instance of the category that does not possess that property. Hence, not one of A to E is a *necessary* property. So, although prototype theories could be thought of as having merely relaxed the classical view's criteria for category membership, the upshot is prototype theory might be able to explain category membership for the many categories that resist definition. (Note the similarities between Table 5.2 and the discussion of diagnosis in Box 5.1 – can you see how the criteria proposed for diagnosing schizophrenia and ME treat these as prototype concepts?)

Prototype theories have been formulated in different ways. Smith *et al.*'s (1988) formalization captures many of the qualities found in different versions. Table 5.3 gives their illustration of a prototype representation for 'apple'.

Table 5.3 highlights some of the differences between prototype theories and the classical view. First, there are multiple possible values for each attribute, capturing the fact that no one value is necessary for category membership – for example,

**Table 5.3** Prototype representation for apple

<b>Diagnosticity</b>	<b>Attribute</b>	<b>Value</b>	<b>Weight</b>
1	COLOUR	red	25
		green	5
		brown	
		...	...
.5	SHAPE	round	15
		square	1
		cylindrical	5
		...	...
.25	TEXTURE	smooth	25
		rough	5
		bumpy	5
		...	...

Source: adapted from Smith et al., 1988

apples are typically, but not necessarily, red. Second, diagnosticities indicate the extent to which each attribute is important for deciding category membership. Third, the values are weighted and these weights indicate the extent to which each value contributes to typicality; the highest weighted values are those of the prototype. Categorization depends upon achieving a criterion similarity with the representation of the concept, one that depends on matching properties as before, but now diagnosticities and weights enter into the computation as well (though we don't need to go into detail). Prototype theories can readily explain the typicality effects discovered by Rosch and her co-workers.

- 1 Instances that differ in typicality are assumed to differ in terms of the weighting of values on which they match the concept. For example, in Table 5.3, a difference in typicality between red and brown apples is reflected in a difference in the weighting for red and brown.
- 2 Sentences such as 'a robin is a bird' are likely to be verified more quickly than 'a penguin is a bird' because, for high typicality instances the criterial similarity required for verifying the sentence is likely to be achieved after matching just a few properties. This is because most attributes that match will have higher-weighted values, and so any criterion for category membership will be reached quickly. For low typicality instances like penguin, many attributes will mismatch or will have low weighted values, and so more matches will have to be made before the criterion is reached.
- 3 Typicality is likely to correlate with how widely category members share attributes. This follows from the fact that the diagnosticities of attributes and weights of values themselves reflect the statistical distribution of those attributes and values. The more widely shared a value is, the greater is its weight. In Table 5.3, for example, 'round shape' receives a high weight

indicating that many (many) more apples are round than square. Since high typicality instances tend to match on high weighted values; it follows that they will also possess properties that are widely shared.

However, despite prototype theory being able to accommodate many of the findings that undermined the classical view, difficulties have emerged, as we shall now see.

### 2.2.1 The meaning of typicality effects

Armstrong *et al.* (1983) considered whether typicality effects occur for concepts that appear to be definitional. Their examples of definitional concepts included ‘female’, ‘plane geometric figure’, ‘odd number’ and ‘even number’ (as in Activity 5.3). Armstrong *et al.* believed that category membership for these concepts is determined not by similarity to a prototype, but by a definition: whether a number is even depends on whether dividing it by 2 yields an integer. Curiously, however, they found a range of robust typicality effects (as in Table 5.4), implying that even these apparently definitional concepts have an internal structure; these effects were also found using the sentence verification task.

**Table 5.4** Typicality ratings for instances of well-defined categories

Even number	Typicality rating	Female	Typicality rating
4	1.1	mother	1.7
8	1.5	housewife	2.4
10	1.7	princess	3.0
18	2.6	waitress	3.2
34	3.4	policewoman	3.9
106	3.9	comedienne	4.5
Odd number	Typicality rating	Plane geometry figure	Typicality rating
3	1.6	square	1.3
7	1.9	triangle	1.5
23	2.4	rectangle	1.9
57	2.6	circle	2.1
501	3.5	trapezoid	3.1
447	3.7	ellipse	3.4

Source: Armstrong *et al.*, 1983

At first glance Armstrong *et al.*'s data could be taken to imply that even concepts such as odd number are not really definitional after all, but organized around a prototype. However, Armstrong *et al.* didn't regard this as a serious possibility. Instead, they argued that the existence of typicality effects should not be taken as conclusive evidence that category membership is determined by similarity to a prototype. They proposed instead a dual-process model, in which concepts possess a 'core' that is used when we judge category membership and a set of identification

procedures that we use to identify instances of a category on particular occasions (often rapidly). Armstrong *et al.* suggested that the classical view might explain the concept's core, while prototype theory explains identification procedures. Unfortunately, inasmuch as this proposal involves both theoretical approaches, it appears to inherit some of the problems faced by each.

### 2.2.2 The context-sensitivity of typicality effects

Another difficulty for prototype theory is the observation that typicality effects change with context. If, as Rosch thought, prototypes reflect natural correlations or clusters of properties, one would expect the prototype to be stable.

However, Roth and Shoben (1983) showed that typicality effects are changed by linguistic context. For example, their participants rated the typicality of different farm animals with respect to the category 'animal'. Participants were first presented with a context sentence that emphasized a particular activity; for example, 'Bertha enjoyed riding the animal' or 'Bertha enjoyed milking the animal'. The context sentence was then followed by a sentence frame such as 'The \_\_\_ quite liked it too'. Participants were asked to rate the typicality of a list of animal words that would complete the sentence frame. Importantly, the list contained words such as 'horse' and 'cow' that fitted well with one context sentence but not with others – though both words were judged to be possible completions of the sentences. Roth and Shoben found that when the context sentence referred to milking, cows were considered to be more typical animals than horses. However, when the context referred to riding, horses were considered more typical animals than cows. (You might notice similarities with the discussion of priming in Chapter 2.)

Medin and Shoben (1988) also found that typicality judgements change with context. They asked their participants to rate various kinds of spoon for typicality in the category 'spoon'. Participants rated metal spoons as more typical than wooden spoons, and small spoons as more typical than large spoons. Therefore, one might expect that small metal spoons would be most typical of all and that large wooden spoons would be least typical, with small wooden and large metal spoons intermediate. However, while Medin and Shoben found that small metal spoons were more typical than large metal spoons, they found that large wooden spoons were more typical than small wooden spoons. So, the contribution to typicality made by the values 'large' and 'small' depended on whether one was thinking about metal spoons or wooden spoons.

Prototype theories cannot easily explain such demonstrations of the instability of typicality. First, the very idea of instability seems to be at odds with Rosch's claim that prototypes correspond to stable clusters of correlated properties that reflect the structure of the natural world. Second, in connection with Table 5.3, Roth and Shoben's results suggest that the weightings of values and/or diagnosticities of attributes are themselves changeable. However, it is unclear what mechanism could be responsible for such changes. Third, Medin and Shoben's results suggest that the contributions to typicality of different properties (e.g. size and material made from) are mutually dependent. Yet the representation in Table 5.3 assumes that the attributes and values are independent of one another.

### 2.2.3 Complex concepts

As noted in Section 1.4, it is commonplace to assume that concepts express our understandings of the meanings of words. So, the concept ‘red’ is assumed to express what we understand as the meaning of the word ‘red’; the concept ‘car’ is thought to provide the meaning of the word ‘car’. But this immediately raises the question: what kind of concept provides the meaning of the phrase ‘red car’?

Researchers have tried to explain the meanings of phrases and larger linguistic units in terms of complex concepts; that is, combinations of lexical concepts. The meaning of the phrase ‘red car’ would then be explained in terms of the combination of the constituent lexical concepts: ‘red’ and ‘car’. How could concepts combine to yield the meaning of such a phrase?

If concepts are structured around prototypes, then perhaps they could combine through combining their prototypes. The difficulty, however, is that no-one really knows how this might be done. Though many suggestions have been made, they all appear to fail for one reason or other. For example, one suggestion has been that the prototype for ‘red car’ is formed from the prototype for ‘red’ and the prototype for ‘car’ (the prototypical red car would therefore be a prototypical car that was prototypically red).

While this seems a sensible suggestion, and appears to give the right interpretation for ‘red car’, this could not work in general. Following the same reasoning, the prototypical ‘pet fish’ ought to be a prototypical fish that is also prototypically pet-like – perhaps something like a cuddly salmon. The real prototypical ‘pet fish’ of course is more like a goldfish – neither a prototypical pet nor a prototypical fish. More problematic still for combining prototypes, the prototypical ‘stone lion’ ought to be something like a real lion made of stone, that is, an impossible object. How could the prototypes for ‘stone’ (perhaps granite or limestone) and ‘lion’ (a real lion) combine to give the right interpretation (i.e. a stone statue of a lion)? If you feel these examples are a little whimsical, take a look at newspaper headlines as these often use phrases with a similar structure. For example, it isn’t easy to see how the meaning of ‘killer firework’ could be explained by the combination of the constituent prototypes: a prototypical killer might be a sadistic criminal, or perhaps a virulent disease; a prototypical firework might be a rocket. How would these prototypes combine to yield the required interpretation? Complex concepts continue to present real difficulties for most theories of concepts (cf. Fodor, 1998).

## 2.3 Common-sense theories: the theory-based view

Both classical and prototype theories explain categorization in terms of similarity using quite simple feature sets. But the problems these theories have encountered have led researchers first to question the importance of similarity and second to propose that categorization involves much larger knowledge structures, called theories (or common-sense theories to distinguish them from scientific ones). The approach has become known as the concepts as theory view or the ‘theory’-theory of concepts.

Before we turn to the ‘theory’-theory we should note, however, that similarity-based accounts have achieved considerable success and remain popular. Hampton (1998) conveys some sense of this. Using McCloskey and Glucksberg’s (1978) data

(they collected both typicality ratings and categorization judgements as you saw in Section 2.1.2), he examined whether the probability of an item being judged a category member could be predicted from its typicality (reflecting its similarity to a prototype).

Focusing on just the borderline cases, Hampton showed that typicality was a very good predictor, explaining somewhere between 46 per cent and 96 per cent of the variance in categorization probability. So, regardless of the difficulties facing similarity-based accounts, similarity (as measured by typicality) seems to be a good indicator of categorization. Nonetheless, Hampton found other predictors of categorization probability (though none was as good a predictor as typicality). These included lack of familiarity; the extent to which an instance was judged ‘only technically speaking a member’ of a category (e.g. a dolphin is technically speaking a mammal, but superficially appears more similar to fish); and the extent to which participants judged an instance was ‘technically speaking not a member’ (e.g. a bat is technically speaking not a bird despite superficially appearing more similar to birds than to mammals). That these last two factors were predictors suggests that categorization draws upon deeper, more theoretical, knowledge than just similarity alone.

We now turn to some of the reasons why, in spite of these successes, many researchers have become dissatisfied with the notion of similarity.

### 2.3.1 Problems with similarity

The philosopher Nelson Goodman identified a number of problems with similarity; indeed, he described it as ‘a pretender, an impostor, a quack’ (Goodman, 1972, p.437). One concern is with whether similarity genuinely helps us to explain categorization. After all, in prototype theories, saying that an instance is similar to the prototype means that the two share some properties in common. But note that this further explication removes the notion of similarity: ‘is similar to’ becomes translated as ‘shares properties with’. So, what explains categorization is not similarity *per se* but the sharing of properties.

However, a further problem arises since there is no obvious limit to the number of properties any two objects may share. Murphy and Medin (1985, p.292) ask us to consider the similarity of plums and lawnmowers: ‘You might say these have little in common, but of course both weigh less than 10,000 kg (and less than 10,001 kg), both did not exist 10,000,000 years ago (and 10,000,001 years ago), both cannot hear well, both can be dropped, both take up space, and so on.’ It seems that, depending on what counts as a *relevant* property, plums and lawnmowers could either be seen as very dissimilar, or very similar. So, for similarity, explicated in terms of shared properties, to provide meaningful explanations of categorization, we need to know what counts as a property. We need some way of declaring ‘lack of hearing ability’ as irrelevant in comparing plums and lawnmowers, for example. For Murphy and Medin (1985), observations such as these suggest that similarity is shorthand for something else that explains why categories hang together, or cohere.

### 2.3.2 The role of common-sense theories

In opposition to similarity-based views, Murphy and Medin argued that concepts are explanation based, that there is some explanatory principle or theory that unites the

category. They offer the example of someone at a party who jumps into a swimming pool fully clothed. You might categorize this person as being intoxicated, but a similarity-based view cannot explain this because your concept of ‘intoxicated’ is unlikely to include the property ‘jumps into swimming pools fully clothed’. So how might we explain the categorization? Murphy and Medin argue that categorizing the person as ‘intoxicated’ plays a role in explaining their behaviour, that is, in explaining why they jumped into the swimming pool.

Might this explanatory basis be found in categorization more generally? If so, then categorizing a robin as a bird ought to provide some kind of explanation of the robin’s properties, analogous to the case of the intoxicated swimmer. Such a categorization does appear to provide a (partial) explanation: knowing that a robin is a bird helps explain why it has feathers and a beak. The explanation is partial, since we could go on to ask why birds have beaks and feathers, but it is an explanation nonetheless. After all, were we to discover that a robin is not a bird, we would want to know why it has feathers and a beak. Without the categorization we would be in need of an explanation.

We noted in Section 2.2.2 that similarity-based approaches cannot easily explain the non-independence of attributes. For Murphy and Medin, relationships between attributes are evidence that our concepts are embedded in larger and broader knowledge structures. Sometimes these structures have been labelled ‘common-sense theories’, sometimes merely ‘background knowledge’. But if such knowledge structures are at work in categorization, why might people have previously concluded that concepts are similarity based? Murphy and Medin speculate that many categorization judgements become automatized, particularly when members of the same category have relatively consistent perceptual properties. Under these conditions, the role of our underlying theories becomes obscured, and so we may (erroneously) conclude that categorization is determined by similarity. However, even in these cases, when novel instances emerge (such as robot cats), or where there is disagreement (with borderlines perhaps), we turn to our underlying theories.

What evidence is there that categorization is determined by theories as opposed to similarity? Rips (1989) asked his participants to consider triads of objects. Two objects belonged to distinct categories (e.g. a pizza and a US quarter) and were chosen so that participants’ largest estimate of the size of one category (the quarter) was smaller than their smallest estimate of the other (the pizza). Rips then asked his participants to consider a third object, telling them only that it was of intermediate size (i.e. larger than the largest estimated size of a quarter and smaller than the smallest estimated size of a pizza). He asked which of the two other categories this third object was more likely to belong to, and which of the two it was most similar to. The two judgements dissociated: that is, participants judged the object more likely to be a pizza, but more similar to a quarter.

Other dissociations between categorization and similarity have been demonstrated (e.g. Rips and Collins, 1993; Roberson *et al.*, 1999). Kroska and Goldstone (1996) showed their participants scenarios that described a putative emotion. Each scenario constituted a set of phrases so that one phrase was central to one emotion and other phrases were characteristic of a different emotion. For example, one scenario included the phrases ‘Threat of harm or death’, ‘Being accepted, belonging’ and ‘Experiencing highly pleasurable stimuli or sensations’. The first of these

phrases was considered central to the emotion category ‘fear’. The remaining two phrases were considered characteristic of the emotion category ‘joy’. Kroska and Goldstone found that their participants tended to categorize this scenario as an instance of fear (i.e. a member of the category ‘fear’) but they also judged it to be more similar to an instance of joy. That is, judgements of category membership were influenced by properties considered central to a category, while judgements of similarity were influenced by characteristic properties. Again, these findings show that judgements of category membership can dissociate from judgements of similarity.

It seems that there are deeper reasons for people’s categorizations – in the quarter example, perhaps they realized that pizzas can, in principle, be any size, whereas their common-sense theories of coins tell them they are produced to a regulation standard (see Box 5.2 for developmental evidence).

## 5.2

### Categorization in development

Support for the idea that knowledge of deeper, causal principles is at work in categorization has come from work looking at children’s categorization. Keil (1989), for instance, used both discovery and transformation procedures to examine how children weigh appearance and theoretical properties. For example, in a discovery, children might be told of a novel hybrid animal that looked and behaved just like a zebra. However, they would be told also that it had been discovered that this animal had the insides of a horse and was the offspring of two horses. Younger children (around 4 years of age) tended to say the animal was a zebra, whereas older children (around 7 years) tended to judge the animal to be a horse. Therefore, younger children seemed to be influenced more by the superficial characteristics of the animal (e.g. appearance), and older children more by its biologically relevant properties (e.g. lineage).

Similar results were found using a transformation procedure. Children were told of a raccoon that underwent a series of transformations so that it ended up looking and behaving like a skunk. For example, it might have skunk-like stripes dyed on its fur, and have a surgical implant so that it could emit foul-smelling liquid. Again, younger children seemed dominated by appearance-based properties; they judged that the raccoon was now a skunk. The older children, in contrast, judged that the animal was still a raccoon.

Keil has referred to this age-related change in children’s categorization as the ‘characteristic-to-defining shift’ since he thought the younger children were influenced by properties (i.e. appearances) that were only characteristic of the category, while the older children were beginning to deploy something like the beginnings of a biological theory, and were paying attention to properties that were more defining. However, as Murphy (2002) points out, it is probably not the case that the younger and older children have qualitatively distinct styles of categorization. It is more likely that the younger children simply do not know enough about biological categories to work out which properties are characteristic, and which are defining.

### 2.3.3 Difficulties with the ‘theory’-theory

The ‘theory’-theory has proved an important and useful way of thinking about concepts. It has, for instance, reminded researchers of difficulties with the notion of similarity, and it has proved to be a useful peg on which to hang a range of disparate findings whose common theme is that categorization is influenced by deeper, causal knowledge of categories, as well as by knowledge of their superficial properties.

However, there are a number of difficulties with the ‘theory’-theory. Some of the findings taken to support the ‘theory’-theory are really demonstrations that similarity does not always explain categorization and this does not necessarily imply that theories are what is needed. Moreover, it is not clear what is meant by ‘theory’. Whereas similarity-based views could be made relatively precise (see Table 5.3 for instance), formalizing ‘theory’-theories seems much more difficult. Some researchers have tried to pin down what is meant by a common-sense theory via a comparison with scientific theories (cf. Gopnik, 1996). However, other researchers believe such a comparison undermines the idea that common-sense theories are theories at all (cf. Gellatly, 1997). For example, Murphy (2000) argues that the background knowledge that influences concepts is too simplistic and mundane to be likened to a scientific theory. Indeed, he eschews the term ‘theory’ in favour of the more neutral ‘knowledge’.

A further difficulty with the ‘theory’-theory is that it is hard to imagine how combining theories could explain complex concepts. Scientific theories are notoriously difficult to combine. Indeed, for decades, theoretical physicists have struggled to combine theories of electricity, magnetism and gravity into one unified theory. So how can theories be combined so effortlessly in understanding phrases like ‘red car’ when they are so difficult to combine in general? Even if we talk of combining knowledge rather than theories, we are still left with the difficult problem of working out which knowledge gets combined and the mechanism by which this is done.

Given these problems, it is ironic that the theory-based view is motivated in part by difficulties with the notion of similarity. Arguably, it has supplanted this with the equally mysterious notion of a ‘theory’.

## 2.4 Psychological essentialism

Psychological essentialism is one attempt at formulating more precisely the view that categorization is influenced by deeper, explanatory principles. Medin (1989) and Medin and Ortony (1989) suggested that people believe that, and act as though, category members have certain essential properties in common. That is, people categorize things according to their beliefs about essential properties. They may also believe that the essential properties constrain a category’s more superficial properties. For example, the essential properties of birds might be thought to involve their genetic make-up, properties that would constrain their appearance and behaviour.

Essential properties can be characterized as properties such that if an object did not possess them, it would not be that object. The essential properties of birds are properties that all birds necessarily possess; if something doesn’t possess them, then it isn’t a bird. Essential properties may seem rather like the defining properties of the classical view. However, there is one critical difference. According to psychological

essentialism most people will not know what a category's essential properties are, but will still believe that the category has some. We might speculate as to what the essential properties are – perhaps for biological categories they would be genetic properties – but, in general, our beliefs will be vague and may turn out to be incorrect. So psychological essentialism proposes that people's concepts may contain a 'place-holder' for an essence – and the place-holder may even be empty, reflecting a lack of knowledge as to what the essential properties might be.

Of course not everyone's place-holder need be empty. Indeed, it is usually thought that discovering essential properties is a job for science. A metallurgist or chemist, perhaps, might uncover the essential properties of gold, just as a biologist might for birds. So, experts may have their place-holders partially or completely filled – they may know (or think they know) the essential properties. But these beliefs may turn out to be in error too, so the place-holder is presumably capable of revision. We can illustrate psychological essentialism with the platypus example of Section 2.1.4. Soon after its discovery, lay-people presumably came to believe that the platypus had a certain essence, but had no idea what this might be (their essence 'place-holder' was empty). Experts at the time might have filled their essence place-holder in different ways: some thought the platypus was essentially an amphibian; others that it was a mammal. But the contents of these place-holders changed as more was learnt. Finally, the experts settled on the view that the platypus was mammalian, and as lay-people adopted this view they filled out their essence place-holder accordingly.

Psychological essentialism is consistent with much of the evidence supporting the 'theory'-theory. Much evidence supporting psychological essentialism specifically has come from studies of the development of categorization (see Box 5.2). For instance, Gelman and Wellman (1991) found that even 4- and 5-year-old children believe the insides of objects to be more important than their outsides in determining category membership. For example, they asked children whether a dog would still be a dog if its outsides were removed, and also if its insides were removed. Children thought that instances would remain in the category if the outsides were removed, but not if their insides were removed. According to Gelman and Wellman, children are being essentialist since they believe that something internal, something hidden and 'inner', is causally responsible for category membership.

However, psychological essentialism has not gone unchallenged. Malt (1994) examined the concept of water. If people believe H<sub>2</sub>O to be the essence of water, then their categorization of liquids as water should be strongly influenced by the proportion of H<sub>2</sub>O those liquids contain. However, Malt found that people's categorizations were strongly influenced by the source of the water, its location and its function. Indeed, pond water was thought to be 'water' but was judged to contain only 78.8 per cent H<sub>2</sub>O; tears were judged not to be 'water' but to contain 88.6 per cent H<sub>2</sub>O. So the belief in the presence or absence of H<sub>2</sub>O was not the only factor in deciding membership in the category 'water'.

In Section 2.1.4, we considered the arguments of the philosophers Kripke (1972) and Putnam (1975). For example, Putnam argued that even if we discovered that all cats are robots controlled from Mars, they would still be cats. What we didn't note there is that they used thought experiments such as this to support essentialism. Braisby *et al.* (1996) subjected these to an empirical test. They asked participants to

give categorization judgements in thought experiments such as Putnam's robot cat. In one condition they were told:

You have a female pet cat named Tibby. For many years people have assumed cats to be mammals. However, scientists have recently discovered that they are all, in fact, robots controlled from Mars. Upon close examination, you discover that Tibby too is a robot, just as the scientists suggest.

Participants were then asked to indicate whether they thought that a series of statements were true or false. These included statements expressing essentialist intuitions (e.g. 'Tibby is a cat, though we were wrong about her being a mammal.') and statements that expressed the contrary intuition (e.g. 'Tibby is not a cat, though she is a robot controlled from Mars.'). Only about half of the participants thought that these essentialist statements were true, and the contrary ones false. Moreover, many participants seemed to give contradictory judgements: they either judged both statements to be true, or judged both to be false. Braisby *et al.* argued that these findings did not support essentialism, but implied that concepts change their content according to context and perspective (cf. Braisby and Franks, 1997).

There has also been mixed evidence concerning the role that expert opinion plays in categorization. Malt (1990) presented people with objects that they were told appeared 'halfway' between two categories (e.g. a tree halfway between an oak and a maple) and asked them to indicate how they would solve the dilemma of categorizing the object. She offered her participants three options. They could 'ask an expert', 'call it whichever you want' or indicate that they could 'tell which it is' if they could only think about it long enough. For pairs of natural categories such as 'robin–sparrow' and 'trout–bass', 75 per cent of participants suggested they would ask an expert, whereas for pairs of artefact categories, such as 'boat–ship', 63 per cent of participants suggested it was possible to 'call it whichever you want'. This evidence suggests that people may be psychologically essentialist for natural categories, at least to some degree, because they recognize that experts may be in a better position to judge categorization when lay-people cannot. However, the data overall are not conclusive. Braisby (2001) examined the extent to which people modify their categorization judgements for genetically-modified biological categories when told the opinions of experts. For example, his participants might be asked to consider a genetically modified salmon, and were told either that expert biologists had judged that it was a salmon or that they had judged that it was not. He found that only around half of the participants changed their categorization judgements to conform to the judgements of the biologists. Moreover, around a quarter of participants would change their categorization judgements to conform to those of shoppers (i.e. a group presumed not to be expert with respect to the category's essential properties). Braisby argued that only around a quarter of participants were modifying their categorization judgements because of the biologists' expertise with the relevant essential properties, and so the majority of responses did not provide evidence for psychological essentialism. Indeed, participants seemed to base their judgements on non-essential properties such as appearance and function (as well as genetic make-up).

Lastly, it should be noted that much of the evidence cited in support of psychological essentialism (e.g. Gelman and Wellman, 1991) only indirectly relates to beliefs in essential properties. Gelman and Wellman, for example, found that children thought that removing the outsides from something like a dog did not alter its category membership, but removing its insides did. However, for these data to support essentialism, a further inference is required, one that relates insides to essences. In a similar vein, Strevens (2000) actually argues that the notion of essence or essential properties is not required to explain empirical data such as these. Of course, psychological essentialists have responded to some of these criticisms so it seems fair to say that the arguments are not yet settled. However, some of the criticisms of other theories may also apply to psychological essentialism – how might it help us understand complex concepts, for example?

## Summary of Section 2

---

- The classical view, that concepts are definitions of categories, is undermined by arguments that many categories cannot be defined, and cannot readily explain typicality effects, borderline cases and intransitivity.
- The prototype view, that categorization is determined by similarity to the prototype, explains most typicality effects. However, it cannot readily explain the context sensitivity of typicality, nor how prototypes might combine in complex concepts. There is a residual question as to whether the existence of typicality effects implies a prototype organization.
- The theory-based view helps explain the non-independence of attributes in concepts, and dissociations between categorization and similarity. It also avoids some of the criticisms aimed at similarity. However, it is not clear how theories might combine in complex concepts, and the notion of a theory is very under-specified.
- Psychological essentialism apparently explains findings that even young children believe inner, hidden properties are causally responsible for category membership. However, it is not clear whether the notion of essence is required to explain data such as these. Moreover, the idea that people categorize according to essential properties has received mixed empirical support, as has the notion that people might defer to expert categorizations.

## 3 Where next?

In this chapter we have canvassed some of the principal approaches that have been taken in developing a theory of concepts. In some respects, it seems as if the study of concepts is the study of theories that do not work for one reason or other. The classical view falters because we cannot identify necessary and sufficient conditions for category membership for all but a very few concepts. Prototype theory has difficulties explaining context sensitivity and complex concepts. Ultimately, both suffer for their use of the notion of similarity, which seems unable to explain

categorization fully. Theory-based notions of concepts are imprecise and cannot obviously explain complex concepts. Lastly, psychological essentialism has received mixed empirical support, and much of the empirical evidence only indirectly relates to the notion of essences.

However, such a picture of the psychology of concepts is unnecessarily gloomy. Indeed, it turns out that we have probably learned more about the phenomena of categorization even as various theories have been found wanting. And, of course, adherents of those theories continue to introduce modifications in order to explain recalcitrant data. Nonetheless, our discussion of the different theoretical approaches raises (at least) two questions. What sense can we make of so many different theoretical treatments, when none is without problems? And where might researchers next turn their attention if there is, as yet, no common theoretical framework? As I shall try to suggest, one way of answering these questions is to consider to what extent categorization is a unitary phenomenon.

### 3.1 Is all categorization the same?

Perhaps the different theoretical treatments of concepts reflect the fact that categorization is not one single process. Maybe people categorize items in different ways in different circumstances. Indeed, discursive psychologists, whose approach we earlier bracketed-off, might argue that categorization depends essentially on context, and that there is nothing common to all the cases that we call categorization. Were context to have such an unbridled influence we might expect categorization to appear unsystematic. Yet, much of the evidence presented in this chapter points to the opposite – we have examined a wide range of empirical data that are highly robust.

One way of reconciling the idea that people categorize things differently on different occasions with the idea that categorization is nonetheless systematic is to suggest that there are (a determinate number of) different kinds of categorization. Moreover, it is conceivable that these could be usefully framed by the different theories of concepts. For example, perhaps the classical view gives a useful account of categorization in cases where we need to provide or appeal to definitions. In law, for instance, often we need to reach an agreement or adopt a convention as to whether something belongs to a category (e.g. whether a 16-year-old is a child or an adult). Similarly, prototype theory may usefully explain categorization in circumstances where we need to categorize something rapidly, or perhaps under uncertainty, maybe when we are in a position to take into account only an object's superficial properties. Likewise, theory-based views may describe categorization when we are seeking a more reflective and considered judgement, perhaps when we are using categorization in order to explain something. And essentialism may usefully explain how we categorize when we wish to be consistent with expertise and a scientific knowledge of the world.

Speculative though this is, Smith and Sloman (1994) have provided evidence that suggests there may be some truth to this possibility. They sought to replicate the dissociation between similarity and categorization judgements obtained by Rips (1989) and described in Section 2.3.2. Rips found that people judged an object intermediate in size between a quarter and a pizza to be more similar to a quarter, but more likely to be a pizza. Smith and Sloman obtained the same dissociation only

when participants were required to think aloud whilst making their decisions and so articulate reasons for their judgements (that is, they provided a concurrent verbal protocol, see Chapter 10). Smith and Sloman interpret this finding as pointing to two modes of categorization: (1) a similarity-based mode of categorization, and (2) a rule-based mode. The implication is that people will either focus on similarity or on underlying rules and structure depending on how the categorization task is presented. When in similarity-based mode, categorization seems to conform to similarity-based accounts, such as prototype theories. When in rule-based mode, categorization seems to be more theory or explanation based. Though this does not show that there are as many different ways of categorizing as there are theories of concepts, it does suggest that categorization may not be a single process. It is a possibility, therefore, that some of the different accounts of concepts may be implicitly concerned with different kinds of purpose in categorization, and ultimately with different kinds of categorization.

In a similar vein we can rethink the phenomena that are taken as evidence of the nature of concepts. Earlier we noted that concepts and words bear a complex relationship to one another, but much of the evidence we have so far reviewed has tended to equate the use of category words with categorization. However, while our use of category labels is certainly influenced by our beliefs about categorization, it is also influenced by language more generally. Indeed, we can label something with a category word yet not believe that it belongs to the category – describing a statue of a lion as a ‘lion’, for example, does not indicate that we think the statue really is a lion. Malt *et al.* (1999) showed how the same is true for how we label containers, such as ‘box’, ‘bottle’ and ‘jar’. They found that whether an item was called a ‘bottle’ depended not so much on how similar it was to a prototypical bottle, but whether there was something similar that was also called a ‘bottle’. In this way, for example, a shampoo container might get called a shampoo ‘bottle’ despite bearing little similarity to a prototypical bottle. So, whether we apply a category label (e.g. bottle) to an object depends in part on how that label has been used historically and only in part on whether we think that the object really belongs to the labelled category (i.e. on whether the object really is a bottle).

### 3.2 Are all concepts the same?

Another possibility that we should consider is the extent to which different types of category require a different theoretical treatment. Already you might have noticed how each theory seems to work most convincingly for a slightly different set of examples. In Activity 5.3 you tried to list the properties of a range of different categories: sparrow, gold, chair, introvert, red, and even number. Did you feel then that these categories were very different from one another? If so, we can perhaps make sense of this intuition.

Some categories like even number seem amenable to definition. For these **well-defined** categories, the classical view appears to give a good explanation of category membership, though it does not obviously explain how some even numbers are considered more typical than others. Perhaps this would require something like Armstrong *et al.*’s dual-process account, and involve its attendant difficulties (see Section 2.2.1). Nonetheless, it may be that a modified classical view would provide a good explanation of these kinds of category.

In a similar vein, prototype theories seem to work well for **fuzzy** categories – like red in Activity 5.3 – categories that seem to have genuine borderline cases. For these, similarity to a prototype might provide the best explanation of category membership, since there is no prospect of defining these categories, nor do people in general seem to have relevant common-sense theories (e.g. a theory of the deeper causal principles by which red things come to appear red). Perhaps categories like chair are fuzzy in the same way.

Theory-based and essentialist approaches are likely to be most successful for categories for which people have common-sense theories. Perhaps unsurprisingly, these include many categories for which scientific theories have also been developed; for example, sparrow and gold, from Activity 5.3. These are categories where it is relevant to develop a deeper, explanatory knowledge of the causal principles underlying the category. Interestingly, it has been argued that essentialism may also help to explain people's concepts of social categories; for instance, introvert in Activity 5.3 (Haslam *et al.*, 2000).

Of course, this is no more than a possibility, and it may be that a single theoretical approach will be devised that can accommodate all of the different kinds of category we have considered. Even if people accepted that different categories require different theoretical treatments, it would still be important to find some way of relating the different theories so we could understand in what sense they were all theories of concepts.

### 3.3 Are all categorizers the same?

Consonant with the above considerations, we might also consider whether all categorizers are the same. Medin *et al.* (1997) recruited participants from three occupational groups with correspondingly different experience and knowledge of trees: maintenance workers, landscapers, and taxonomists. They then asked them to sort the names of 48 different kinds of tree into whatever groups made sense. The taxonomists tended to reproduce a scientific way of sorting the trees; the maintenance workers produced a similar sorting, although they gave more emphasis to superficial characteristics (such as whether trees were broad-leaved). They also tended to include a 'weed tree' group that was not present in the taxonomists' sorts, and which included trees that cause particular maintenance problems. The landscapers didn't reproduce a scientific taxonomy, but justified their sorts in terms of factors such as landscape utility, size and aesthetic value. Lynch *et al.* (2000) also showed how the typicality ratings of the same kinds of tree expert differed from those of novices. Typicality for the expert group reflected similarity to ideals, so trees judged to be best examples of the category were not of average or prototypical height, but of extreme height; in contrast, the ratings of the novices were largely influenced by familiarity.

Studies such as these suggest that different people do not necessarily categorize things in the same way. The goals that a person has as well as the extent of their knowledge may influence the way they categorize and, by extension, be reflected in their concepts.

## Summary of Section 3

It is possible that the failings of one or all of the approaches to concepts may be due to any combination of the following:

- Categorization may not be a single process; and different kinds of categorization may lend themselves to different theoretical treatments.
- Different types of category have different properties and so may require different theoretical treatments.
- Different groups of people may categorize things in different ways, according to their goals and the nature of their knowledge, and so may fit the claims of different theories.

## 4 Conclusion

Overall, it seems that category knowledge is multi-layered, encompassing knowledge of the causal properties relevant to a category, knowledge relevant to explaining category membership and the properties of instances, knowledge of function, and knowledge of superficial properties useful for identification and judgements about appearance. It also seems that we are capable of calling on different kinds of category knowledge on different occasions and for different purposes. While these observations are not inconsistent with a single theoretical treatment of concepts, they nonetheless raise the prospect that competing theories provide good explanations of somewhat different sets of phenomena, and so are not directly in contradiction. However the theoretical debates may or may not be resolved, I hope this chapter has convinced you of the importance of concepts to an understanding of cognition. Though categorization presents substantial challenges for researchers, these are challenges for all cognitive psychologists. Only once they have been met are we likely to be able to develop a good understanding of the mind. (None of which is likely to trouble Rosie.)



**Figure 5.2** Rosie (untroubled)

## Answer to Activity 5.1

Here are the identities of the objects shown in Figure 5.1 (from left to right): olive stoner; asparagus peeler; pickle picker, ideal for retrieving the very last pickled onion or gherkin from a jar.

## Further reading

Inevitably in a chapter of this length, I have omitted some important issues. Most notably, I have not touched on the exemplar view of concepts, the literature on category learning, or the issue of basic level concepts. For these, I would strongly recommend Greg Murphy's excellent book. For a philosophically inspired selection of psychological and philosophical works, see Laurence and Margolis.

Laurence, S. and Margolis, E. (1999) (eds) *Concepts: Core Readings*, Cambridge, MA, MIT Press.

Murphy, G.L. (2002) *The Big Book of Concepts*, Cambridge, MA, MIT Press.

## References

- American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4th edn), Washington, DC, American Psychiatric Association.
- Armstrong, S.L., Gleitman, L. and Gleitman, H. (1983) 'What some concepts might not be', *Cognition*, vol.13, pp.263–308.
- Barsalou, L.W. (1983) 'Ad hoc categories', *Memory and Cognition*, vol.11, pp.211–27.
- Barsalou, L.W. (1985) 'Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.11, pp.629–54.
- Braisby, N.R. (2001) 'Deference in categorization: evidence for essentialism?', in Moore, J. D. and Stenning, K. (eds) *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Mahwah, NJ., Lawrence Erlbaum.
- Braisby, N.R. and Franks, B. (1997) 'What does word use tell us about conceptual content?', *Psychology of Language and Communication*, vol.1, no.2, pp.5–16.
- Braisby, N.R., Franks, B. and Hampton, J.A. (1996) 'Essentialism, word use and concepts', *Cognition*, vol.59, pp.247–74.
- Bruner, J.S., Goodnow, J.J. and Austin, G.A. (1956) *A Study of Thinking*, New York, John Wiley.
- Chater, N. and Heyes, C.M. (1994) 'Animal concepts: content and discontent', *Mind and Language*, vol.9, pp.209–46.
- Eco, U. (1999) *Kant and the Platypus: Essays on Language and Cognition*, London, Secker & Warburg.
- Edwards, D. and Potter, J. (1992) *Discursive Psychology*, London, Sage.
- Fodor, J.A. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford, Clarendon Press.

- Gellatly, A.R.H. (1997) 'Why the young child has neither a theory of mind nor a theory of anything else', *Human Development*, vol.40, pp.1–19.
- Gelman, S. and Wellman, H. (1991) 'Insides and essences: early understandings of the non-obvious', *Cognition*, vol.38, pp.213–44.
- Goodman, N. (1972) 'Seven strictures on similarity', in Goodman, N. (ed.) *Problems and Projects*, Indianapolis, IN, Bobbs-Merrill.
- Gopnik, A. (1996) 'The scientist as child', *Philosophy of Science*, vol.63, pp.485–514.
- Hampton, J.A. (1982) 'A demonstration of intransitivity in natural categories', *Cognition*, vol.12, pp.151–64.
- Hampton, J.A. (1998) 'Similarity-based categorization and fuzziness of natural categories', *Cognition*, vol.65, pp.137–65.
- Haslam, N., Rothschild, L., and Ernst, D. (2000) 'Essentialist beliefs about social categories', *British Journal of Social Psychology*, vol.39, pp.113–27.
- Hull, C.L. (1920) 'Quantitative aspects of the evolution of concepts', *Psychological Monographs*, vol.28.
- Keil, F. (1989) 'Concepts, kinds and cognitive development', Cambridge, MA, MIT Press.
- Kripke, S.A. (1972) 'Naming and necessity', in Davidson, D. and Harman, G. (eds) *Semantics of Natural Languages*, Dordrecht, Reidel.
- Kroska, A. and Goldstone, R.L. (1996) 'Dissociations in the similarity and categorization of emotions', *Cognition and Emotion*, vol.10, no.1, pp.27–45.
- Lynch, E.B., Coley, J.D. and Medin, D.L. (2000) 'Tall is typical: central tendency, ideal dimensions and graded category structure among tree experts and novices', *Memory and Cognition*, vol.28, no.1, pp.41–50.
- Macintyre, A. (1998) *ME: Chronic Fatigue Syndrome: A Practical Guide*, London, Thorsons.
- Malt, B.C. (1990) 'Features and beliefs in the mental representation of categories', *Journal of Memory and Language*, vol.29, pp.289–315.
- Malt, B.C. (1994) 'Water is not H<sub>2</sub>O', *Cognitive Psychology*, vol.27, pp.41–70.
- Malt, B.C., Sloman, S.A., Gennari, S., Shi, M. and Wang, Y. (1999) 'Knowing versus naming: similarity and the linguistic categorization of artefacts', *Journal of Memory and Language*, vol.40, pp.230–62.
- McCloskey, M. and Glucksberg, S. (1978) 'Natural categories: well-defined or fuzzy sets?' *Memory and Cognition*, vol.6, pp.462–72.
- Medin, D.L. (1989) 'Concepts and conceptual structure', *American Psychologist*, vol.44, no.12, pp.1469–81.
- Medin, D.L., Lynch, E.B., Coley, J.D. and Atran, S. (1997) 'Categorization and reasoning among tree experts: do all roads lead to Rome?', *Cognitive Psychology*, vol.32, pp.49–96.
- Medin, D.L. and Ortony, A. (1989) 'Psychological essentialism', in Vosniadou, S. and Ortony, A. (eds) *Similarity and Analogical Reasoning*, Cambridge, Cambridge University Press.

- Medin, D.L. and Shoben, E.J. (1988) 'Context and structure in conceptual combination', *Cognitive Psychology*, vol.20, pp.158–90.
- Murphy, G.L. (2000) 'Explanatory concepts', in Keil, F.C. and Wilson, R.A. (eds) *Explanation and Cognition*, Cambridge, MA, MIT Press.
- Murphy, G.L. (2002) *The Big Book of Concepts*, Cambridge, MA, MIT Press.
- Murphy, G.L. and Medin, D.L. (1985) 'The role of theories in conceptual coherence', *Psychological Review*, vol.92, pp.289–316.
- Pinker, S. (1997) *How the Mind Works*, London, Penguin Books.
- Potter, J. and Wetherell, M. (1987) *Discourse and Social Psychology*, London, Sage.
- Putnam, H. (1975) 'The meaning of "meaning"', in *Mind, Language and Reality*, vol. 2, Philosophical papers, Cambridge, Cambridge University Press.
- Rips, L.J. (1989) 'Similarity, typicality and categorization', in Vosniadou, S. and Ortony, A. (eds) *Similarity and Analogical Reasoning*, Cambridge, Cambridge University Press.
- Rips, L.J., Shoben, E.J. and Smith, E.E. (1973) 'Semantic distance and the verification of semantic relations', *Journal of Verbal Learning and Verbal Behaviour*, vol.12, pp.1–20.
- Rips, L.J. and Collins, A. (1993) 'Categories and resemblance', *Journal of Experimental Psychology: General*, vol.122, pp.468–86.
- Roberson, D., Davidoff, J. and Braisby, N. (1999) 'Similarity and categorization: neuropsychological evidence for a dissociation in explicit categorization tasks', *Cognition*, vol.71, pp.1–42.
- Rosch, E.H. (1973) 'On the internal structure of perceptual and semantic categories', in Moore, T.E. (ed.) *Cognitive Development and the Acquisition of Language*, New York, Academic Press.
- Rosch, E.H. (1975) 'Cognitive representations of semantic categories', *Journal of Experimental Psychology: General*, vol.104, pp.192–233.
- Rosch, E.H. and Mervis, C.B. (1975) 'Family resemblances: studies in the internal structure of categories', *Cognitive Psychology*, vol.7, pp.573–605.
- Ross, B.H., and Murphy, G.L. (1999) 'Food for thought: cross-classification and category organization in a complex real-world domain', *Cognitive Psychology*, vol.38, pp.495–553.
- Roth, E.M. and Shoben, E.J. (1983) 'The effect of context on the structure of categories', *Cognitive Psychology*, vol.15, pp.346–78.
- Sappington, B.F. and Goldman, L. (1994) 'Discrimination learning and concept formation in the Arabian horse', *Journal of Animal Science*, vol.72, no.12, pp.3080–7.
- Smith, E.E., Osherson, D.N., Rips L.J. and Keane, M. (1988) 'Combining prototypes: a selective modification model', *Cognitive Science*, vol.12, pp.485–527.
- Smith, E.E. and Sloman, S.A. (1994) 'Similarity- versus rule-based categorization', *Memory and Cognition*, vol.22, no.4, pp.377–86.
- Strevens, M. (2000) 'The essentialist aspect of naive theories', *Cognition*, no.74, pp.149–75.

Sutcliffe, J.P. (1993) ‘In defence of the “classical view”: a realist account of class and concept’, in van Mechelen, J., Hampton, J.A., Michalski, R. and Theuns, P. (eds) *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, London, Academic Press.

Wittgenstein, L. (1953) *Philosophical Investigations* (trans. by G.E.M. Anscombe), Oxford, Basil Blackwell.



# Language processing Chapter 6

Gareth Gaskell

---

## 1 Introduction

What are the qualities of human beings that differentiate us from other species? You can probably think of many characteristics, but pretty high on most people's lists would be the ability to produce and understand language. Linguistic abilities underpin all manner of social interactions – from simple acts such as buying a bus ticket or greeting a friend, right up to constructing and refining political and legal systems. Like many aspects of cognition, the ability to use language develops apparently effortlessly in the early years of life, and can be applied rapidly and automatically.

This chapter looks under the surface of the language system, in order to understand the unconscious operations that take place during language processing. Our focus is on the basic mechanisms required for language understanding. For example, understanding a simple spoken sentence involves a whole string of abilities: the perceptual system must be able to identify speech sounds, locate word boundaries in sentences, recognize words, access their meanings, and then integrate the word meanings into a coherent whole, respecting the grammatical role each word plays. Each of these abilities has been extensively researched, with numerous models of how information is processed being proposed and tested, and this chapter provides an overview of our current understanding in these cases. As you will see, there often remains considerable disagreement about some quite fundamental properties of the language system. Nonetheless, there has also been substantial progress in terms of identifying some of the features required of the language system for it to work the way it does.

The building blocks of language identified in this chapter are discussed in a wider context in Chapter 7, which examines, for example, questions such as how speakers interact in conversation. Chapter 7 also covers language production, whereas the current chapter concentrates on language *perception*. The structure of this chapter roughly follows the time course of processing in language perception. Section 2 builds on some of the ideas about recognition introduced in Chapter 4, but looks specifically at the processes that result in the identification of spoken and written words. Models of these processes generally assume that word recognition involves access to a **mental lexicon** – something that was briefly introduced in Chapter 5 in the context of lexical concepts – which stores relevant information relating to the words we know (e.g. what they mean). Section 3 deals with the contents of the mental lexicon, and how this information might be organized. Finally, Section 4 looks at the process of sentence comprehension beyond the mental lexicon. It deals with how listeners use their knowledge of the grammar of a language to construct the meaning of a sentence. In each section some of the influential models of language processing are discussed, along with key experimental studies that help us to evaluate and refine these models.

## 2 Word recognition

Adult speakers of English tend to know somewhere between 50,000 and 100,000 words. Most common words are easy to describe and use, suggesting their meanings are clearly accessible. Less common words are perhaps represented more vaguely, with some words difficult to define out of context, but nonetheless generating a feeling of familiarity. For example, you might be reasonably confident that *tarantella* is a word and have good knowledge of how it should be pronounced, but you might still be unable to give a good definition of what it is (a fast whirling dance, once believed to be a cure for a tarantula bite!).

So quite a lot of information is stored in the mental lexicon about word meanings and pronunciations. The goal of word recognition is to access this information as quickly as possible. We shall look at how this process occurs in two different sensory modalities: auditory and visual. This may at first seem repetitious, but there are some important differences between the two modalities that, at this level of the language system, lead to quite different models of recognition processes. Before you read through the sections on word recognition, you may wish to remind yourself of the broader issues involved in recognition, as described in Chapter 4.

### 2.1 Spoken word recognition

Speech is the primary medium of language. Widespread literacy has emerged only in some cultures, and only in the last century or two, meaning that reading is, in evolutionary terms, a new ability. Speech in contrast is something that almost all humans acquire, and has been around long enough for some aspects of spoken language to be thought of as innate. Speech is also primary in the sense that we learn to understand and produce speech before we learn to read and write. For these reasons, we will firstly look at how spoken word recognition operates, and then go on to examine the visual modality.

#### 2.1.1 Segmenting the speech stream

#### ACTIVITY 6.1

---



**Figure 6.1** A speech waveform

The waveform in Figure 6.1 depicts a typical sound wave that might enter the ear when you hear someone speak. Try to work out from the sound wave how many words have been spoken, and pencil in a mark where you think each word boundary lies.

### COMMENT

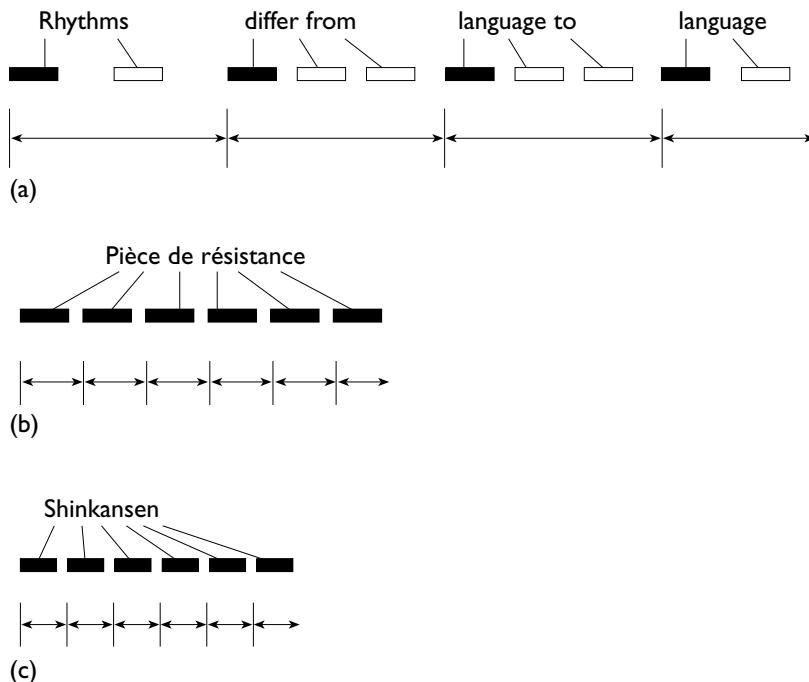
When you have noted down your estimates, compare them with the actual boundaries marked at the end of the chapter. How did you decide on likely word boundaries, and was this method a useful one? Most people assume that silent gaps between words are likely boundary markers, but they can be misleading. Some word boundaries do not involve silence because the surrounding **phonemes** are **coarticulated**, meaning that they blend together. A phoneme is the speech equivalent of a letter (they are normally annotated with surrounding slash marks), so, for example, /k/ and /ə/ are the first two phonemes in *confess*. Coarticulation refers to the fact that you have to prepare for upcoming phonemes well before they are produced, and these preparations lead to changes in the phonemes currently being pronounced. For example, the /d/ phoneme in 'do' and 'dah' sounds slightly different because of the following vowel. In addition, some silent gaps do not mark word boundaries: they are just points where the airways are closed in the course of uttering a word. For example, when you say the word 'spoken' your lips close briefly in order to produce the sudden release of air in the phoneme /p/. This results in a short period of silence between the /s/ and the /p/.

---

Our conscious experience of spoken words is in some ways similar to our experience of text on a page: words are perceived as coherent and discrete events, so we generally don't experience any difficulty in finding the dividing line between two words. However, the truth of the matter is that the speech waveform has no simple equivalent of the white space between printed words. Instead, as Activity 6.1 shows, silent gaps are unreliable as indicators of spoken word boundaries. Yet somehow the language system must be able to divide the speech stream up, so that the words contained in it can be recognized and understood. How then does this **word-segmentation** process operate?

Models of segmentation can generally be divided into two types: (1) **pre-lexical** models and (2) **lexical** models. Pre-lexical models rely on characteristics of the speech stream that might mark a likely word boundary, whereas in lexical models segmentation is guided by knowledge of *how words sound*. The first model is pretty straightforward: the only issue at stake is what type of characteristic or cue can be extracted from the speech waveform as a useful indicator of a word boundary. We have already seen that silent gaps are not sufficient, but nonetheless silence can be useful, particularly if it lasts quite a long time.

Another important pre-lexical cue comes from the *rhythm* of speech. All languages have some unit of temporal regularity, and this provides the basic rhythm when an utterance is produced. In English, this unit is known as a **metrical foot**, and consists of a strong (stressed) syllable, followed optionally by one or more weak (unstressed) syllables (as you can see from Figure 6.2). Strong syllables are



**Figure 6.2** Examples of language rhythms. In English (a) the basic unit of rhythm is the strong syllable (the filled boxes). These stressed syllables are roughly equally spaced out in time when you produce a sentence, no matter how many weak syllables (unfilled boxes) there are between the strong syllables. Each group of strong and weak syllables is known as a ‘foot’, so when you say the sentence in (a), it may feel like you are speaking more quickly towards the end of the sentence because you need to fit in more weak syllables to maintain the gaps between the strong syllables. In French (b), the syllable is the unit of rhythm and so all syllables are roughly equally spaced in time. The rhythmic unit in Japanese (c) can be even smaller than a syllable. For example, *shinkansen* (‘bullet train’) contains six units of rhythm (including three single consonants), but only three syllables

reasonably clear landmarks in the speech stream, and most words that have a meaning (such as *bacon* or *throw*) rather than a grammatical role (e.g. *it*, *of*) begin with a strong syllable (Cutler and Carter, 1987). So, a segmentation strategy that predicts a word boundary before each strong syllable would seem like a valuable one for English speakers.

Cutler and Norris (1988) provided evidence supporting this idea: they played pairs of nonsense syllables to listeners, and asked them to monitor for any familiar word embedded in the speech (this is known as the **wordspotting** task – think trainspotting but duller). For example, in the sequence ‘mintayve’, which consists of two strong syllables, there is the embedded word ‘mint’. Cutler and Norris argued that for a sequence like this listeners should identify the two strong syllable onsets (the /m/ and the /t/), and search for any words they know beginning at those points. This segmentation would obscure recognition of the word ‘mint’, because it spans a hypothesized word boundary (i.e. they would tend to hear two units: ‘min’ and ‘tayve’). On the other hand, a sequence like ‘mintesh’ (where the second weak

syllable contains a reduced ‘uh’ vowel) would tend to be segmented as a single unit, and so spotting the word ‘mint’ should be relatively easy. Their prediction turned out to be correct, suggesting that listeners make use of the rhythm of English in order to identify likely word boundaries. In languages where different rhythmic units dominate, such as French (syllables) or Japanese (sub-syllabic units), similar sensitivities have been demonstrated (see Cutler and Otake, 2002), suggesting that early in life people ‘tune into’ their native language and optimize their segmentation strategy accordingly.

These (and other) pre-lexical cues are clearly valuable for identifying likely word boundaries in a sentence of utterances. However, none of the models that rely on pre-lexical cues can claim complete accuracy in boundary identification. This means that there will be cases where a boundary is incorrectly predicted, and other cases where a real boundary is missed. For example, a word like *confess* begins with a weak syllable, and so its onset would be missed by a pre-lexical segmentation strategy based on strong syllables. It seems that there must be some other mechanism available for cases like this. This is where lexical models can offer more insight: lexical segmentation models rely on our knowledge of particular words’ **phonological representation** (what they sound like) to guide segmentation. The simplest version of this kind of strategy would involve recognizing each word in an utterance sequentially, and so predicting a new word at the boundary of the existing word (e.g. Marslen-Wilson and Welsh, 1978). For example, think about how the sentence ‘Confess tomorrow or die!’ might be segmented. If you can recognize the first word quickly (before it finishes), then you can use the knowledge that this word ends in /s/ to predict a word boundary as soon as the /s/ is encountered. You can then start again on word recognition with the speech following the /s/ (*tomorrow*). The problem here though is that most words are much shorter than *confess* and *tomorrow*, and cannot be recognized within the time-span of their acoustic waveforms, meaning that a lot of backtracking would be required to locate word boundaries using this method (think about trying to segment the sentence ‘Own up now or die’ using the same strategy).

We shall return to this issue in Section 2.1.3, when we evaluate the TRACE model of spoken-word recognition (McClelland and Elman, 1986), which provides a more powerful lexical-segmentation mechanism. Although there is plenty of evidence supporting pre-lexical mechanisms, it remains likely that lexical competition operates alongside them to provide a more robust system for dividing up the speech stream.

**6.1****Research study****Learning to segment speech**

The cross-linguistic differences between segmentation mechanisms highlight the fact that the ability to segment speech is one that must be learnt during the course of language development. French and English babies aren't innately specified with different segmentation mechanisms; instead these develop as a consequence of exposure to language. Saffran *et al.* (1996) provided an impressive demonstration of how statistical information can aid the development of both segmentation and vocabulary acquisition.

They devised an artificial language made up of three-syllable words such as *dapiku* or *golatu*, and then used a computer to synthesize a long continuous stream of speech containing these 'words'. Their intention was to produce a sequence in their artificial language that contained absolutely no acoustic or rhythmic cues to the location of word boundaries. If people only make use of acoustic and rhythmic segmentation cues then the speech they hear should appear as unsegmented nonsense. However, if they can make use of statistical information about co-occurrence of syllables, then they may start to pick out the words of the language. In other words, they might start to notice that the syllables *da*, *pi* and *ku* quite often occur in sequence.

Using what is known as a 'head-turning' procedure, 8-month-old infants were tested on their perception of this kind of speech. The infants were presented with words from the artificial language on one loudspeaker and jumbled syllables (e.g. *pikugo*) on another. The idea was that if the infants found the words from the language familiar, they might spend more time listening to the novel sequences (and turn their heads towards the associated loudspeaker). Using this technique, Saffran *et al.* (1996) found that the infants *did* begin to pick out the words from the stream of syllables after just two minutes of the speech. This ability to learn the statistical properties of patterns is quite universal – it operates for adults and children as well as babies, and works just as well for nonspeech stimuli such as tones or shapes (Saffran *et al.*, 1999). Therefore, speech segmentation may make use of a wide-ranging **implicit learning** ability, which may even be shared by other primates, such as tamarin monkeys (Hauser *et al.*, 2001).

**2.1.2 Parallel activation**

A spoken word typically lasts about half a second. In many ways it might simplify matters if the recognition process began only once the whole of a word had been heard. However, for the language system, this would be valuable time wasted. Instead, speech is continually evaluated and re-evaluated against numerous potential candidates for the identity of each word: this is known as **parallel activation**. A great advantage of this method of assessment is that it can lead to determination of a word's identity well before the end of the word is heard.

The mechanism sketched above is most clearly exemplified by the cohort model of Marslen-Wilson and colleagues (Marslen-Wilson and Welsh, 1978; Marslen-Wilson, 1987). This model assumes that as the beginning of a word is encountered, the **word-initial cohort** (a set of words that match the speech so far) is activated. For

example, if the beginning of the word were ‘cuh’ (as in *confess*), then the word-initial cohort would include words like *canoe*, *cocoon*, *karate* and so on, because these words all match the speech so far. Then, as more of the word was heard, the recognition process simply becomes one of whittling down the set of potential candidates. For example, ‘conf...’ would rule out all the words above, but not *confess*, *confetti*, or *confide*. At some point in this process (the **uniqueness point**) the candidate set should be reduced to a single word. According to the cohort model, the recognition process is then complete. As mentioned above, the recognition point in this kind of model can be well before the end of the word, meaning that valuable time is saved in interpreting the speaker’s message.

However, even this conception of the process doesn’t reflect the full fluency of word recognition. So far, we haven’t discussed the *goal* of the recognition process – accessing our stored knowledge about a word. One might assume that this occurs at the recognition point of a word. However, it seems that access to meaning can occur substantially earlier. Marslen-Wilson (1987) demonstrated this using **cross-modal priming**. This technique – which is used to examine the extent to which the meaning of a spoken word has been retrieved – involves hearing a spoken **prime** word, followed swiftly by a visual **target** word. Participants were given the task of deciding whether the target was a word or not as quickly as possible. Semantic similarity between a prime-target pair such as ‘*confess*’ and *sin* leads to faster responses to the target (compared with an unrelated control pair, such as ‘*tennis*’ and *sin*). This implies that, on reaching the end of the word, the meaning of ‘*confess*’ has been activated. The question that Marslen-Wilson addressed was whether the meaning would be activated at an earlier point, before the uniqueness point had been reached. He found that when something like ‘*confe...*’ was used as a prime, responses to the target word *sin* were still facilitated. The same spoken fragment would also facilitate responses to the target *wedding*, which was semantically related to an alternative cohort member, *confetti*. This suggests that the meanings of both *confess* and *confetti* are briefly accessed while the word *confess* is being heard.

You might want to reflect on what this result means in terms of how we recognize spoken words. It suggests that when we hear a word, we don’t just activate the meaning of that word, we also activate, very briefly, the meanings of other words that begin with the same phonemes. Meanings of likely candidates are activated before the perceptual system can identify the word being heard, which ensures that the relevant meaning has been retrieved by the time the word is identified.

Parallel activation of multiple meanings is an important property of the language recognition system. The alternative – a serial search, which would be a bit like looking through a dictionary for a word meaning – is unlikely to be as efficient (particularly for words near the end of the list). However, it is worth questioning the extent of parallel activation. For example, could it be the case that there is no limit to the number of meanings that can be activated briefly? And can these multiple meanings be accessed without any interference between them? Gaskell and Marslen-Wilson (2002) argued that meaning activation is limited, again on the basis of cross-modal priming data. They showed that if many meanings are activated at the same time, the resultant priming effect is relatively weak compared with the amount of priming found when just one or two meanings are compatible with the speech input. It appears to be the case that activating more than one meaning can only occur

partially, so the gradual reduction of the cohort set of matching words is accompanied by a gradual isolation and amplification of the relevant meaning. Nonetheless, an overriding characteristic of the speech perception system is to access *too much* information rather than too little. This maximizes the chances of having accessed the correct meaning as soon as enough information has been perceived to identify the particular word.

### 2.1.3 Lexical competition

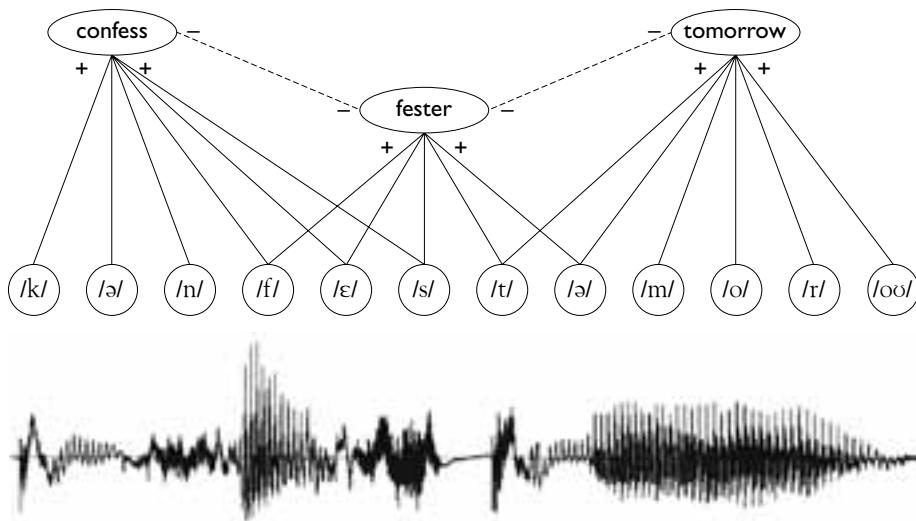
Marslen-Wilson's cohort model was important because it incorporated parallel evaluation of multiple lexical candidates, and emphasized the swiftness and efficiency of the recognition process. Later models used a slightly different characterization, and relied on the activation and competition metaphor (introduced in Chapter 4 in the discussion of the IAC model of face recognition). In these models, each word in the lexicon is associated with an **activation level** during word recognition, which reflects the strength of evidence in favour of that particular word. The cohort model in its original form (Marslen-Wilson and Welsh, 1978) can be thought of as a dichotomous activation model: words are either members of the cohort (equivalent to an activation level of 1) or they aren't (activation level 0). The advantage of more general models of lexical competition such as TRACE (McClelland and Elman, 1986) is that they can use continuously varying activation levels to reflect the strengths of hypotheses more generally. This is useful in cases where a number of words are consistent with the speech input so far, but the information in the speech stream matches some words better than others. If activation levels are on a continuous scale, then this inequality can be reflected in the activations assigned to word candidates.

The TRACE model is a connectionist model that assumes three levels of representation: the **phonetic feature** level (phonetic features are basically bits of phonemes), the phoneme level, and the word level (containing a node for each word the listener knows). The idea of the model is that the speech stream is represented as changing patterns of activation at the phonetic feature level. These nodes feed into a phoneme recognition level, where a phonemic representation of speech is constructed. A word node has connections from all the phonemes within that word. For example, the *confess* node would have connections feeding into it from the /k/, /ə/, /n/, /f/, /ɛ/, and /s/ phoneme nodes. If the phoneme nodes for that word became activated, activation would then spread to the *confess* word node, resulting in strong activation for that word. The net result is that word-node activations reflect the degree to which each word matches the incoming speech.

A second mechanism provides a way of selecting between active words. Nodes at the word level in TRACE are connected by inhibitory links. When any word node becomes activated, it starts to inhibit all other word nodes (i.e. by decreasing their activation) with the strength of inhibition depending on the degree to which that node is activated. This competitive element tends to amplify differences in word activations, so that it becomes clear which words are actually in the speech stream and which are just *similar to* the words in the speech stream. So if the spoken word was *confess*, then the node for *confetti* would become strongly activated as well, because all phonemes in the input apart from the final one fit the representation of *confetti*. However, the *confess* node would be activated to a slightly greater extent

because all phonemes in the input are consistent. Both these word nodes would be strongly inhibited by the other, but the greater **bottom-up support** (i.e. greater consistency with the incoming signal) for *confess* would ensure that the *confess* node would eventually win the competition, remaining activated when the *confetti* node had been strongly inhibited.

This ‘winner-takes-all’ activation and competition approach is common to many models both within language (we shall see another example in Section 2.2) and across cognition (e.g. face recognition). These commonalities across different areas of cognition are valuable, as they provide a way of extracting more general principles of cognitive processing from specific examples. In the case of speech, lexical competition provides a simple mechanism for deciding which words best match the speech input. As we saw in Section 2.1.1, it also provides a subsidiary means of segmenting the speech stream into words. This is because it is not just words which have the same onset, such as *confess* and *confetti*, that compete, but also words that simply overlap to some extent, such as *confess* and *feſter* (see Figure 6.3). These words have a syllable in common (i.e. the second syllable of *confess* and the first syllable of *feſter*). If this syllable is perceived, then both of these words will become activated, but through lexical competition only one will remain active. The segmentation problem can then be viewed as having been solved *implicitly* in the



**Figure 6.3** Illustration of lexical competition in the TRACE model. The speech stream activates a set of phonetic feature nodes (not shown), which then activate the corresponding phoneme nodes. Word nodes at the lexical level are linked up to the relevant phoneme nodes with positive links (solid lines). In this case, the speech is actually ‘confess tomorrow’. This sequence actually fits three words completely: *confess*, *feſter* and *tomorrow*. These word nodes have inhibitory connections between them that vary in strength depending on their degree of overlap. So there is no inhibitory link between the *confess* and *tomorrow* nodes, but the *feſter* node has inhibitory links to both *confess* and *tomorrow* (broken lines). The combined inhibition of the *feſter* node from the other nodes has the effect of suppressing its activation, leaving only *confess* and *tomorrow* as active lexical candidates. The competitive links between words in TRACE allow word recognition to be carried out, and also provide a mechanism for word boundary identification (there must be a boundary between the /s/ of *confess* and the /t/ of *tomorrow*)

activation of the word nodes. For example, if *confess* wins the competition then there must be a word boundary at the end of the syllable ‘*fess*’, but if *feſter* wins the competition then the boundary must be at the start of ‘*fess*’.

This general version of lexical competition is supported by a wordspotting experiment by McQueen *et al.* (1994). They looked at the time taken to spot a word like *mess* in two different types of embedding sequence. In a sequence like ‘duhmess’ the first two syllables match a longer word: *domestic*. If lexical competition operates for all overlapping words (see Figure 6.3), then the inhibitory link from the *domestic* node should make it difficult to spot *mess*. McQueen *et al.* found that detection rates were indeed lower and slower in this case, as compared to a case like ‘nuhmess’, in which a longer competitor does not exist. Lexical competition appears to be a rather neat way of performing two essential processes (word identification and segmentation) at the same time. Word identification performed in this way has the added bonus of providing a partial solution to the segmentation problem.

## 2.2 Visual word recognition

In this section we focus on the special qualities of word recognition in the visual domain, looking at how the recognition process operates, how visual and auditory processes are linked, and how eye movements are linked to the recognition system. Compared to speech, text might be thought of as an unproblematic medium. After all, it is relatively easy to spot where words begin and end, and text isn’t transient in the way that speech is – if you misperceive a word on the page, you can simply go back to that word and try again. However, the availability of textual information also raises specific issues that must be addressed by models of visual word recognition. For example, because textual information is freely available over an extended period of time, we need to understand how the recognition system determines where the eyes should fixate, and for how long.

### 2.2.1 Models of visual word recognition

We have already seen how TRACE models *spoken* word recognition in terms of activation and competition in a multi-level connectionist network. TRACE was in fact a variant of an earlier model of visual word recognition proposed by McClelland and Rumelhart (1981). The visual model is often known as the IAC (interactive activation and competition) model, and shares many properties with the IAC model of face recognition you met in Chapter 4. The model contains three levels of nodes, representing activation of (1) visual features, (2) letters and (3) words. Like TRACE, it is the inhibitory units within a level that provide a competitive activation system. Visual input is represented by activation at the **featural** level, and facilitatory and inhibitory links between levels of representation allow activation to build up at the higher levels. In this way, visual word recognition can be modelled as an interactive competition process.

An important property of many of these competition networks is that as well as allowing activation to flow up through the system (i.e. from features through letters to words), during the course of recognition they also allow activation to flow in the other direction (from words downwards). This is another example of the concept of top-down processing that was introduced in Chapter 3. For example, if the word

node for *slim* became activated, the activation would feed back through facilitatory links to the constituent letter nodes (i.e. ‘s’, ‘l’, ‘i’ and ‘m’). At first glance, these feedback links appear redundant, because the letter nodes are going to be activated in any case by bottom-up sensory information. But their value becomes apparent in cases where the bottom-up information is degraded in some way. For example, suppose that the first letter of *slim* was obscured slightly, so that the ‘s’ letter node was only weakly activated by the visual input. In this case, the *slim* word node would still be activated by the three unambiguous letters, and would in turn increase the activation of the ‘s’ letter node. The result would be correct recognition of the obscured letter, despite the weakness of the sensory evidence.

This kind of top-down influence can be useful in explaining **lexical effects** on lower-level processing. A classic finding in word recognition (known as the ‘word superiority effect’, or WSE) is that letter detection is easier when the letter forms part of a word (e.g. the letter ‘i’ is easier to detect in *slim* than in *spim*). This can be attributed to the influence of the word node for *slim* providing a secondary source of activation for recognition of ‘i’, whereas there is no secondary source for a non-word like *spim*. So the top-down feedback connections in the IAC model provide a neat explanation of why we often find lexical influences on recognition of sublexical units like letters.

However, Grainger and Jacobs (1994) demonstrated that a variant of the IAC model could also explain the WSE without any top-down feedback. They proposed that responses to the letter-detection task were based on two different levels of representation: a letter-detection response could be based on activation of letter nodes *or* word nodes. The idea here was that one of the pieces of information about a word stored in the mental lexicon is a description of the written form of the word. So if a word node reaches a critical level then the spelling of that word should be activated, triggering a response. The upshot was that the incorporation of a second basis for responses using lexical information allowed the WSE to be accommodated in a model that only used bottom-up flow of activation.

The experimental finding of WSE remains a robust and important phenomenon, but the research of Grainger and Jacobs shows that there is more than one way of explaining the effect. Whether or not top-down processing is needed is one of the most contentious questions in the area of word recognition (both auditory and visual) and other areas of perception, and it remains a hotly debated topic amongst cognitive psychologists (e.g. Norris *et al.*, 2000, and associated commentaries).

### 2.2.2 Mappings between spelling and sound

So far we have treated the question of how words are recognized separately for spoken and written words. This section looks at how these two modalities interact, and what this tells us about the language system. There is an obvious need for interaction in order to spell a word that you have just heard, or read aloud in a written sentence. But there are more subtle reasons for suspecting that there are links between the **orthography** of a word (its spelling) and its **phonology** (its sound). Some of the data we shall now look at suggest that visual word recognition relies strongly on spoken word representations and processes.

## ACTIVITY 6.2

---

Think about what processes might operate when you read aloud the following words: *bell*, *stick*, *pint*, *yacht*, *colonel*. How does a reader convert the orthographic (or written) form to a phonological one in order to pronounce the words, and where is the phonological information stored? Would the same processes operate when you read the following non-words: *dobe*, *leck*, *brane*, *noyz*?

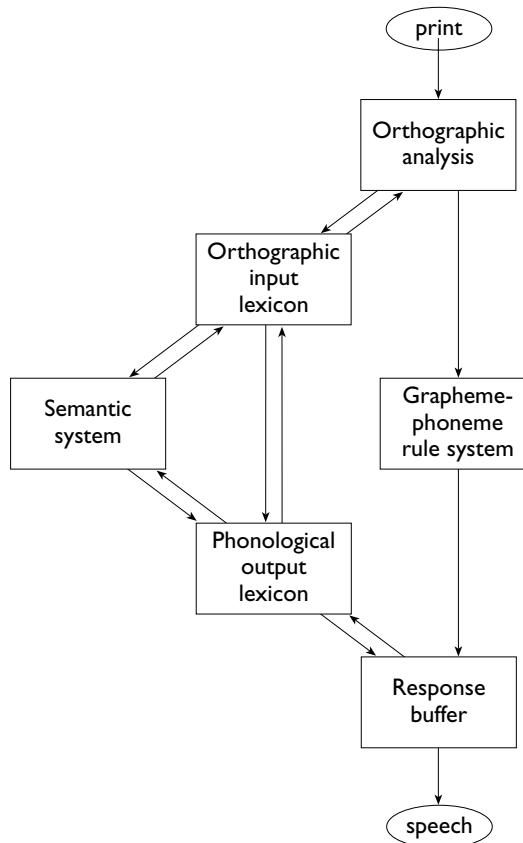
### COMMENT

Researchers often refer to two different ways of reading words aloud (what we could call 'retrieving their phonology'). The division is much like the division between phonics and whole-word methods of teaching children to read. **Assembled phonology** (like phonics) means generating a pronunciation based on a set of mappings between letters and sounds for your language. For example the 'b' in *bell* corresponds to the /b/ phoneme, and there are similar conversion rules for 'e' and 'll'. This works well for words such as *bell* and *stick*, because they follow these conversion rules (i.e. they are regular items), and also for non-words like *dobe* and *leck*. But what about *pint*, *yacht* and *colonel*? A simple sounding out of these irregular words would lead to the wrong pronunciation (e.g. *pint* might be pronounced to rhyme with *hint*), suggesting that an alternative mechanism is available. This is often known as **addressed phonology**, and (like whole-word methods of teaching reading) relies on some kind of stored pronunciation of the whole word in the mental lexicon. *Brane* and *noyz* are unusual because their pronunciation coincides with the pronunciations of real words (i.e. *brain* and *noise*). These **pseudohomophones** (non-words that can be pronounced to sound like words) are generally only found in rock lyrics and some rather fiendish language experiments (see below).

---

As described in Activity 6.2, reading aloud is often portrayed in terms of two separate mechanisms: assembled and addressed phonology. The separate mechanisms are explicitly represented in dual-route models of reading such as the DRC model of Coltheart *et al.* (2001). DRC (see Figure 6.4) is a complex and powerful model, and builds on more than 100 years of theorizing about multiple routes in reading processes. For current purposes, the critical feature of the model is that it contains a 'rule-based' route to pronunciation via a grapheme-phoneme rule system (assembled phonology; see right-hand side of Figure 6.4), plus a 'lexical' route that requires retrieval of a stored pronunciation (addressed phonology; see left-hand side of Figure 6.4). Looking at the speed with which written words can be named often assesses the degree to which these routes are involved in reading. A typical finding is that regular words are named faster than irregular words (e.g. *pint*), but that this advantage is only present for **low-frequency** words (i.e. words that occur relatively rarely in the language). This can be explained by dual-route models in terms of a race between the two routes. Regular words can be named via either the lexical or the rule-based route to pronunciation, whereas irregular words can only make use of the lexical route. On the whole, naming speeds are faster when two routes are available (naming a regular word can be based on the output of whichever route delivers the pronunciation first), than when only one route is available (irregular words). For

low-frequency words, the advantage of two routes over one for regular words results in them being named more quickly. For high-frequency words, it is assumed that the lexical route operates very quickly regardless of regularity, and so the influence of the additional rule-based route is minimal.



**Figure 6.4** DRC model

Source: Coltheart et al., 2001, Figure 6

Regularity is not the only variable that determines the speed with which a word can be named. Glushko (1979) showed that the properties of **neighbouring** words – words that have similar spellings, *not* words that are in neighbouring locations – are also critical. For example, people are quick to name a word like *wade*, because it is a regular word, but also because all neighbouring words with the same final letters have a consistent pronunciation (e.g. *made*, *jade*, *spade*). On the other hand, although *wave* is also a regular word, its neighbours are inconsistent in terms of pronunciation (e.g. *have* and *slave* don't rhyme). This inconsistency results in slower naming. Simple dual-route models, which apply the same rules to regular items irrespective of their neighbours, could not easily explain consistency effects, and an alternative conception of the spelling-sound mapping arose partly as a response. Seidenberg and McClelland (1989) proposed that a single connectionist network could provide a basis for modelling naming of both regular and irregular items, while accounting for effects of neighbouring items, as in Glushko's consistency effect.

There is an obvious need for phonology to be accessed in reading aloud, because speaking requires a phonological representation. But does phonology also have a role to play when a reader simply has to identify and understand written words? Van Orden (1987) showed that phonological representations are involved in silent reading even when they are detrimental to performance. Van Orden asked participants to decide whether visually presented words were members of particular categories, such as whether a *rose* is a flower. Critically, participants found it difficult to reject **homophones** to a category member, such as *rows*. In these cases participants would frequently make an incorrect response, suggesting that they were activating the pronunciation of the homophone, and this was creating confusion. A similar effect was found when the critical items were pseudohomophones (e.g. *roze*).

Other demonstrations have consolidated the idea that spoken word representations are heavily involved in visual word recognition in many different languages. This may seem rather bizarre – surely word recognition based on visual features would be simpler and quicker? But we need to remember that speech perception is to some extent an innate ability, and we learn to understand spoken language very early in life. So when we begin to read, we already have a perfectly tuned recognition system for speech. It therefore makes sense for the visual recognition system to ‘latch onto’ the spoken system in order to ease the learning process. A major issue however relates to whether and how the phonological system can be bypassed later in life as reading becomes more skilled (Frost, 1998).

### 2.2.3 Eye movements in reading

Speech perception is a relatively passive process, in that the listener doesn’t need to perform any overt action in order to listen to a conversation. Reading a book or newspaper, however, is more active, because the reader controls the speed of uptake of information, and must direct their eyes in order to take in new information. Eye movements turn out to be enormously useful in revealing how the language system operates.

As discussed in Chapter 3, eye movements may feel quite smooth and continuous introspectively, but they really consist of **saccades** (jerky movements), followed by **fixations** (more-or-less stationary periods) during which visual information is processed. Eye-tracking techniques can monitor the movements of the eyes during reading, and relate them to the location of the reader’s gaze (see Box 6.2). Figure 6.5 illustrates the fixations involved in the processing of a typical sentence. Each numbered circle corresponds to the gaze location for a single fixation. Fixations typically last about 200 ms, but their durations are strongly dependent on the linguistic processing involved. For example, fixation duration is strongly dependent on the frequency of a word’s usage in the language (Rayner and Duffy, 1986). This, along with many other effects, suggests that fixations are a measure of some kind of processing difficulty, and so they can reveal influential variables in reading.

## 6.2

## Methods

### Eye tracking

Eye-tracking techniques generally rely on the fact that various parts of the eye such as the lens and the cornea reflect light. If a light source (usually infrared) is directed at the eye from a given angle, the angle of the reflection can be used to determine the orientation of the eye, and consequently the direction of gaze. Precise measurements can be made if the eye-tracking system combines measurements from more than one surface within the eye.

In studies of reading, the position of the head is often fixed using a chinrest and headrest and the participant is presented with text on a computer screen. Given that the head position is fixed and the distance from the screen is known, the reader's gaze location relative to the text can be calculated from the gaze-angle measurements. This results in a set of timed fixations to the text, as illustrated in Figure 6.5.

It is clear how eye-tracking studies would be beneficial for understanding how we read. However, a less obvious use of eye tracking is in the study of spoken language. Here, the participant is presented with a spoken sentence in the context of some visual scene, and the eye movements of the listener are monitored. For example, if a participant is sitting in front of a table with some candy and a candle on it, and is asked to pick up the candy, it is revealing to find out at what point people look at the candle, and correlate this with the amount of speech information they have received at that point (Tanenhaus et al., 1995). In this kind of situation (see Section 4.4 for another example of this method), the participant needs to be able to move their head freely. To allow for this, a slightly different type of tracker is used, consisting of an eye tracker mounted on the head plus a second system for determining head position.

As Figure 6.5 illustrates, our eyes don't simply move from one word to the next as reading proceeds. Some words are skipped altogether, whereas others require multiple fixations. In a significant proportion of cases, readers perform regressive saccades (i.e. they move backwards through the text), as marked by the grey circle in Figure 6.5. Short **function words** (grammatical words like *we* and *on*) are much more likely to be skipped than **content words** (words that convey meaning, like *sentence* and *look*), and regressions can often tell us about cases where a word has been misinterpreted, due to some ambiguity (Starr and Rayner, 2001).

Where<sup>1</sup> we<sup>2</sup> look<sup>3</sup> when<sup>4</sup> reading<sup>5</sup> a<sup>6</sup> sentence<sup>7</sup> is<sup>8</sup> dependent<sup>9</sup> on<sup>10</sup> many<sup>11</sup> different<sup>12</sup> factors<sup>13</sup>.

**Figure 6.5** Example of typical eye movements during reading

Eye-movement data are also valuable in terms of understanding where we fixate *within* a word. O'Regan and Jacobs (1992) showed that words are identified most quickly if they are fixated at a point in the word known as the **optimal viewing position**, or OVP. The OVP is generally near the middle of a word, but can be

slightly left of centre in the case of longer words. The fact that fixations work best if they are near the middle of the word makes sense, given that visual acuity is best in the **foveal** (central) region of the retina (try fixating on the edge of this page and reading the text!). This slight but consistent bias in favour of left of centre is more intriguing. Shillcock *et al.* (2000) argued that this bias reflects a balancing of the informativeness of the parts of the word to the left and right of the fixation point. The OVP should be left of centre for longer words because there is greater redundancy towards the end of most of these words. For the word *cognition*, for example, it would be easier to guess what the word is from the first five letters (*cogni*) than the last five (*ition*). Shillcock *et al.* also found that for some shorter words such as *it*, the theoretical OVP was outside the word, either to the left or the right, perhaps explaining why shorter words are often not fixated when reading.

```
xxxxx wxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx (3)
*  

xxxxxxxxxxxxxxxxxxxxding xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx (5)
*  

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxtence depxxxxxxxxxxxxxxxxxxxxx (9)
*  

xxxxxxxxxxok when readingxxxxxxxxxxxxxxxxxxxxxxxxxxxxx (15)
*  

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxentence depends on many dxxxxxxxxxxxxx (25)
*
```

**Figure 6.6** Example stimuli in a moving window experiment (e.g. McConkie and Rayner, 1975). The numbers in parentheses are the window sizes in characters. In a typical experiment the participant sits in front of a computer screen, with an eye-tracking system monitoring her gaze. Wherever the participant directs her gaze along a line of xs the computer displays around that point small 'windows' of unobscured text. All text outside this window is obscured by xs (in some experiments the gaps between words were preserved). The asterisk below each example marks the fixation point, and would not be seen by a participant (participants can fixate where they like). In this case the sentence is 'Where we look when reading a sentence depends on many different factors'

How much textual information can be utilized when a reader makes a fixation? The foveal region of the retina has the greatest acuity, but it spans a limited angle. It is possible that information can also be gained from the **parafocal** region, which is wider but has reduced acuity. Rayner and colleagues have carried out a number of ingenious experiments aimed at assessing the perceptual span of readers. In one experiment they applied a moving window to text shown on a computer screen so that only a certain number of letters to the left and right of the current fixation point could be read (see Figure 6.6). Whenever the participant shifted their gaze, the window shifted accordingly. They found that if the window is small, reading is a slow and painful process, but for larger windows participants are barely aware of the text beyond the window that is masked.

The idea behind this technique is that the experimenter can gradually increase the window size until a point is reached at which reading speed and comprehension are normal. At this point one can be confident that the text beyond the window range is not used for normal reading. The results suggest that the perceptual span for English readers is quite limited: about 15 characters to the right of fixation and 3 characters to the left. The asymmetry is due to the left-to-right nature of reading in English – there

is more useful information to be gained in text following the fixation position than the text preceding the fixation (which has normally already been read). In a right-to-left language like Hebrew the asymmetry is swapped, showing that the perceptual span in reading is shaped by the requirements of the written language.

## Summary of Section 2

---

- In speech, finding out where words begin and end is a nontrivial problem. Models of word segmentation rely on either features of the speech stream or knowledge about how words sound.
- Word recognition relies on parallel assessment of multiple options, and competition between word candidates.
- Models of word recognition differ in the extent to which top-down processing is required.
- Visual word recognition relies to a considerable extent on speech codes.
- Studying eye movements during reading reveals what aspects of visual word recognition cause difficulties.

## 3 The mental lexicon

In Section 2, word recognition was largely viewed as an identification procedure. That is to say, we assumed that the mental lexicon stores representations of what words sound and look like, and that when we hear or see a word there is a recognition process that compares the input with stored representations. However, identification is just the first step towards understanding a word – what we *really* need to know is what a word means. In this section we shall move beyond identification, and look more closely at how word meanings are accessed during the recognition process. We shall look at the **semantic content** (how word meanings are stored) and the **semantic organization** (how word meanings are related) of the mental lexicon. Before we do this, we shall also take a brief look at quite a difficult area of language processing known as **morphology**.

### 3.1 Morphology

Morphology deals with the size of units in the mental lexicon. It's often taken for granted that the basic unit of the mental lexicon is the word, but in fact many words can be broken down into **morphemes** (the smallest meaningful unit within a word) when they are perceived. This implies that the morpheme is the true basic unit. This is particularly the case for languages such as Turkish in which words tend to be rather long, cumbersome units, and a lexicon based on morphemes would be more economical (fewer entries) and more flexible.

Most people are aware that words can be divided up into meaningful units, and that these units follow some grammatical rules. For example, we know that plurals in English are generally derived by adding an ‘s’ to the singular form, as in *cats* or *dogs*. In its spoken form, the rule is slightly more complex: speakers add on /s/, /z/ or /iz/,

depending on the final phoneme of the singular form (think about how you would say the plural forms *cats*, *dogs* and *pieces*). This kind of relatively minor modification of a word (for example, marking pluralization or tense) is known as an **inflectional change**, and is covered by a branch of morphology known as **inflectional morphology**. More major modifications are possible as well, in which the grammatical class of a word may change. For example, the suffix *-ness* can change an adjective to a noun (as in *happiness* or *weakness*). Similarly, the suffix *-ly* can change an adjective into an adverb. These modifications form part of **derivational morphology**.

The examples of morphological change given above are particularly straightforward. All involve regular changes in which the meaning of the word is predictable from the meanings of the morphemes. But things are not always so simple. For example, according to the regular pluralization rule, the plural form of *mouse* should be *mouses* not *mice*. *Mice* is an example of an irregular plural form, and similar irregularities exist in many types of morphological change. Similarly, the meanings of the morphemes making up a word may not always determine the meaning of the whole word. It's easy to spot the relationship in meaning between *govern* and *government*, but not between *depart* and *department*, yet both pairs have (at least supposedly) the same morphological relationship.

The descriptions given here are linguistic ones, but do they have any relevance to cognitive psychology? In other words, do the regularities that exist between families of words have any effect on the organization of the mental lexicon? It is quite possible that our recognition system is set up to recognize words, regardless of their substructure. This **full-listing** approach would mean that recognizing a word made up of many morphemes such as *disenchantment* is essentially the same process as recognizing a single-morpheme word such as *cat*. The opposite extreme – often known as the affix-stripping or **decompositional** approach (Taft and Forster, 1975) – is that words are chopped up into morphemes as they are perceived, and the morpheme is the basic unit of representation in the lexicon.

One way of looking at whether the lexicon is organized in terms of morphemes is to test whether we can add morphemes onto unknown words. A classic demonstration of this kind of generalization is Berko's 'wug test' (1958), in which children were encouraged to generate the plural form of novel words. For example, a child might see a drawing of a toy and be told that it was a *wug*. The child would then see two of these toys and be prompted to say what they were. Children found it easy to generate the correct inflected form (*wugs*), suggesting that they had learnt some kind of pluralization rule, and that the 's' can operate as an independent morpheme. The pluralization suffix is a particularly common one, but other morphemes such as *-ment*, as in *government* or *en-*, as in *enact* are more rare. This factor may affect the way different morphemes are represented in the lexicon. It may be that common morphemes such as the plural morpheme are stored as separate units, but less common units have no separate representation.

Marslen-Wilson *et al.* (1994) used the priming methodology to examine whether morphemic units exist in the mental lexicon. Their reasoning was that if words are broken down into morphemes then we should be able to get strong priming effects between words containing the same morpheme. They found that priming of this type depended on some shared meaning between the two words. So hearing *cruelty*

resulted in faster processing of *cruel* because they have similar meanings, but hearing *casualty* did not prime responses to *casual*, presumably because there was no clear link between the meanings of the two words.

These results suggest that extreme positions such as full-listing or full decomposition are untenable. Factors such as the transparency of the semantic link between morphemes and words determine the extent to which morphemes are represented in the mental lexicon. So it may make more sense to have a pragmatic view of morphological processing, in which morphological decomposition only occurs if there is some clear benefit to be had. Different morphemes within a language may be treated in different ways, and there may also be differences between languages in terms of the extent to which the mental lexicon relies on morphemes.

## 3.2 Accessing word meanings

Chapter 5 introduced you to the notion of lexical concepts – a class of concepts specific to words. In this section we shall relate the ideas underlying concepts and categories to the operation of the semantic system. We shall examine the kinds of information that become available once a word has been recognized, and also look at the problem of how to select the appropriate meaning in cases where a word is ambiguous.

### 3.2.1 Semantic representations

Once a word has been recognized the relevant information about that word must be accessed, so that the word, and ultimately the sentence, can be understood. Most models of language perception start to get slightly hazy at this point, because while word forms are quite concrete and easy to define, their meanings are rather less tangible, and may vary quite strongly from person to person.

Two theories of how word meanings might be represented have gained popularity since the 1970s, both of which have links to the kinds of ideas discussed in Section 2.2 with respect to interactive activation models. Spreading activation models (e.g. Collins and Loftus, 1975) assume that words can be represented by units or nodes, as in the TRACE and IAC models of word recognition. The difference here is that links between nodes in spreading activation networks represent semantic relationships between words. Collins and Loftus's original model in 1975 used different kinds of links for different kinds of semantic relationship. For example, the network could encode the fact that a canary is a bird, by linking the nodes for *canary* and *bird* with an 'is a' link, or that a canary has wings using a 'has' connection. Other models didn't use labelled links but simply connected together words that were similar in meaning. The application to word recognition would be that once a word has been recognized (for example by activating the correct node in the IAC model), activation would spread to the semantic network, and then along links to related words, thus generating a set of known facts about that word, and activating a set of semantically related words.

The alternative **featural** theory of semantic representation assumes that word meanings are represented as a set of **semantic features** or properties (a bit like some of the theories of concepts explored in Chapter 5). The idea here is that the mental

lexicon contains a large set of features, and that each word representation consists of a subset of these features. For example, the features relevant for the word *canary* might include ('has wings', 'can fly', 'is a bird' and so on). The feature model has also been incorporated into connectionist models of recognition, allowing the linkage of recognition models and semantic representations. In this case, the activation of a written or spoken representation would lead to a pattern of activation on a set of semantic nodes, with each node representing a semantic feature (e.g. Masson, 1995).

These two approaches are highly underspecified, and could potentially accommodate many different patterns of data. Despite this, you might find it useful when you read through the experimental findings listed below to think about how the findings might be accommodated by featural and spreading activation theories. Most studies of semantic representations of words have addressed what kinds of information can be accessed and when. Clearly, all kinds of information about a word could be stored in the mental lexicon, but the information required to understand a sentence must be readily available in a fraction of a second, and this time constraint may have some consequences for what types of information are stored.

The most popular tool for investigating the types of semantic information stored in the mental lexicon is semantic priming. For example, an experiment might use pairs of semantically related words, such as *bread* and *butter*, with participants asked to perform some kind of speeded task such as lexical decision (is it a word or not?) or naming (say the word aloud) to the second item of the pair. In this case, the assumption is that if responses are facilitated (i.e. quicker) when there is a semantic relationship between the words, then that semantic relationship must be represented in the mental lexicon (in a spreading activation model there might be a link between the words).

So what kinds of relationship between words can support semantic priming? Perhaps the most robust effect involves pairs of **associated** words (words that seem to go together naturally). **Association strength** is often measured by asking people to say or write down the word that first comes into their heads when they read a target word. So if you were asked to provide an associate for the word *cheddar* you would probably say *cheese*. According to the University of South Florida norms (Nelson *et al.*, 1998), that's what more than 90 per cent of respondents say (curiously, a further 3 per cent of their respondents said *Swiss*!). In any case, the fact that presenting one word results in facilitated processing of an associated word suggests that associative links between words are represented in the lexicon in some way.

The problem with this conclusion is that the types of relationship found for associated word pairs are quite variable, ranging from near synonyms (words that have very similar meanings, such as *portion* and *part*) to antonyms (words that have opposite meanings, such as *gain* and *lose*), to words that just crop up in the same context (e.g. *law* and *break*). For this reason, researchers have often tried to look for semantic priming in cases where words have only weak associations, but still retain some specific semantic link (e.g. *horse* and *sheep*). The data here are more equivocal, which suggests that non-associative links might be weaker in some way, or rely on a different mechanism compared with associative links.

Nonetheless, Lucas (2000) reviewed a large set of semantic priming experiments and reached the conclusion that non-associative semantic priming effects were robust, with perhaps the strongest evidence for links in the lexicon between members of the same category (e.g. *horse–pig*) and instrument–action pairs such as *broom* and *sweep*.

Kellenbach *et al.* (2000) looked at whether words might be linked in terms of the visual or perceptual properties of the objects they represented. For example, *button* and *coin* both refer to flat, round objects. This kind of priming had been observed weakly in some studies, but not others. However, Kellenbach *et al.* (2000) used two measures of priming: the first was the standard reaction time test, and the second was based on brain activity using the ERP technique (see Box 6.3 in Section 3.2.2). They found no effect in the reaction time test, but nonetheless a robust effect on the brain response to the target word, suggesting that even in this case, where the semantic link was too subtle to be detected by conventional techniques, a priming relationship still existed. So it seems that the semantic information that becomes available when a word is perceived is far from minimal. Instead, many different aspects of meaning are accessed. Current research says little about how these different aspects of meaning are organized and accessed, but even at this stage it seems that associative, pure semantic, and perceptual knowledge might be accessed in different ways.

### 3.2.2 Semantic ambiguity

In many cases, the operation of activating a word's meaning in the mental lexicon is made more difficult because the word is ambiguous in some way. For example, what does the word *bank* mean to you? You may immediately think of a high-street bank, but then later realize that *bank* could mean the side of a river as well. This is because *bank* is a homonym: a word that has multiple unrelated meanings. There are also more subtle possibilities: the first meaning of *bank* is most commonly applied to the place you keep your money. But a blood bank, while clearly related, is a somewhat different concept, as is the bank at a casino. So *bank* is a polysemous word, as well as a homonym, because it has multiple related senses. Further ambiguity is caused by the fact that *bank* could be a verb (transitive or intransitive) or a noun, but we shall leave this **syntactic ambiguity** to the next section. Homonyms are thankfully reasonably rare (roughly 7 per cent of common English words according to Rodd *et al.*, 2002), but the vast majority of words have multiple senses, which means that we really need to deal with ambiguity effectively if we are going to understand language.

Normally, the sentential context of an ambiguous word will provide some valuable clues to allow the relevant meaning of the word to be selected. So the question that researchers have focused on is *how* sentential context influences meaning selection in cases of ambiguity. Two opposing views have emerged since the 1980s (you may note similarities between the debate here and the debate on top-down and bottom-up processing discussed in Section 2.2). According to the **autonomous** view, all meanings of an ambiguous word are first accessed, and then the contextually compatible meaning is selected from these alternatives. The **interactive** view has a stronger role for sentential context, in that it may in some

cases rule out inappropriate meanings before they are fully accessed. So these two viewpoints differ in terms of whether there is a short period of time in which meanings of words are accessed regardless of sentential context.

Using cross-modal semantic priming, Swinney (1979) found evidence for autonomous activation of ambiguous word meanings. In his experiment, participants heard homonyms like ‘bugs’ embedded in sentential contexts, and were asked to make a lexical decision to a visual target related to one of the meanings of the prime or an unrelated control word (see Figure 6.7).

### UNBIASED CONTEXT

**Hear:** ‘Rumour had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several bugs in the corner of his room’

**See:**

ANT / SPY / SEW



### BIASED CONTEXT

**Hear:** ‘Rumour had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches and other bugs in the corner of his room’

**See:**

ANT / SPY / SEW



**Figure 6.7** Example trial in Swinney’s (1979) priming experiment. In the unbiased context, both meanings of bugs are plausible (relating to insects and relating to spying). The activation of each meaning is assessed using the reaction time to a related word (ant or spy), compared with a control unrelated word (sew). In the biased context, only the insect meaning is plausible by the time the homonym is heard

Swinney found that whether or not the sentence context was biased towards one meaning of the homonym, both related targets were primed. This implies that both meanings of the ambiguous word were accessed, despite the fact that in the biased condition only one meaning was compatible with the sentential context. When the experiment was repeated with the targets presented roughly one second later, only the contextually appropriate meaning appeared to be activated. So Swinney’s data suggested that there is a short window of up to a second in which the meanings of ambiguous words are accessed without regard to sentential context, supporting the autonomous model.

Variants of Swinney’s experiment have been run many times, and once again there is some inconsistency in the pattern of priming. In some cases it seems that only one meaning is activated if the homonym has one particularly common meaning and the sentential context is strongly constraining towards that meaning (Tabossi and Zardon, 1993). Lucas (1999) has also shown that studies demonstrating exhaustive access of ambiguous word meanings often still show *more* priming for the contextually appropriate meaning than the inappropriate one. Therefore it seems that at least some interactive processing is likely in accessing word meanings, although sentential context may only rule out inappropriate meanings in specific circumstances.

**6.3****Methods****Event-related potential (ERP) studies of semantic processing**

The ERP methodology relies on the fact that brain activity creates an electromagnetic field that can be measured by a set of electrodes placed on the scalp. Typically, the recording of activity is synchronized with the presentation of a stimulus, and many recordings using different stimuli must be averaged to generate an interpretable waveform. The resultant ERP waveform often contains a set of characteristic peaks at different delays.

A negative peak occurring roughly 400 ms after the stimulus has been presented (known as the N400) has been identified with the integration of semantic information into sentential context. A typical finding is that the size of the N400 peak associated with a word in sentential context is inversely related to how easily that word fits into the context (Kutas and Hillyard, 1980). So the N400 peak associated with the word *spoon* might be small in the sentence 'James ate the cereal with a dessert spoon', but large in the sentence 'James caught the salmon using a fishing spoon'. This sensitivity to semantic congruity makes the ERP technique an excellent one for examining issues such as lexical ambiguity resolution.

Van Petten and Kutas (1987) compared ERP and standard priming methods of assessing the effects of sentential context on meaning activation for ambiguous words such as *bank*. They showed that even when standard priming techniques detected no influence of sentential context the ERP waveforms for the ambiguous words were subtly different, suggesting that sentential context was influencing the processing of these words, and strengthening the case for an interactive account of lexical ambiguity resolution.

---

## Summary of Section 3

---

- The mental lexicon stores the meanings of words. Although the subject is contentious, it seems that some words are broken down into smaller units called morphemes.
- A wide variety of information about the meaning of a word becomes available when a word is recognized, including associative knowledge, pure semantic information and perceptual features.
- For words with more than one meaning, the sentential context of the meaning can help select the relevant meaning. This process is to some extent interactive.

## 4 Sentence comprehension

So far, language perception has largely been described in terms of recognition processes. Up to the level of the lexicon, the job of the perceptual system is simply to allow recognition of familiar sequences (words or morphemes) and retrieve stored

knowledge relating to these items. When we discussed morphology, there was a little more **productivity** involved. That is, people can recognize and make use of novel morphological variants of familiar morphemes. So, for example, even if the word *polysemous* were new to you when it was mentioned in the previous section, you would probably find it quite easy to define its morphological relative, *Polysemy*. However, when we get to the level of the sentence, the character of language perception changes abruptly. Sentences are almost always new, in that the same permutation of words has often never been encountered before. If perception at this level were still simply a recognition process, then we would completely fail to understand all but the most simple or common sentences. The solution to this problem is to treat sentence-processing not as a pure recognition process but as a *constructive* process. When we read or hear a sentence, we take the individual components – the words – and combine them to produce something that may be quite novel to us, but hopefully bears some relationship to the message the speaker or writer intended. You might think of this process in terms of building up a mental model of the information being communicated (see Chapter 12 on reasoning). Accordingly, the listener or reader takes each word and deduces its grammatical or syntactic role in the current sentence. Termed **parsing**, this process is the focus of the final section in this chapter.

## 4.1 Syntax

### ACTIVITY 6.3

---

Please read the following passages and sentences and think about whether they seem grammatical to you. Give each one a rating from 1 to 10, where grammatical sounding passages get high marks.

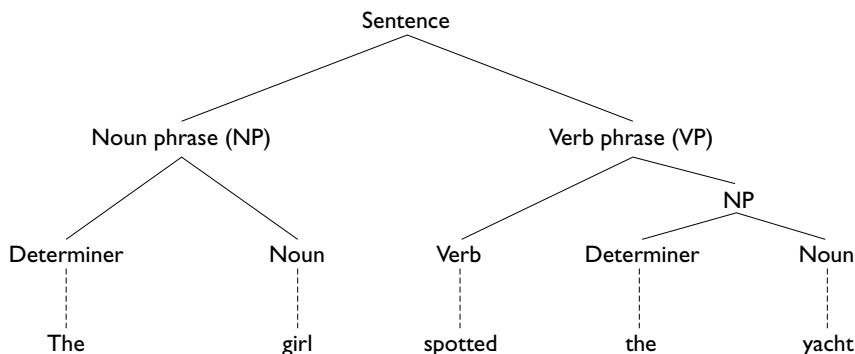
- 1 The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.
- 2 Her five-year mission: to explore strange, new worlds; to seek out new life and new civilizations; to boldly go where no man has gone before.
- 3 Please cup, gimme cup.
- 4 Colourless green ideas sleep furiously.
- 5 In become words sentence the rather have jumbled this.

### COMMENT

People have quite reliable intuitions about the grammaticality of sentences, despite often being unable to define exactly what makes a sentence grammatical. You probably gave the first two passages fairly high ratings. Passage 1 is a quotation from Albert Einstein, and applies quite well to the study of syntax: mysterious but potentially very revealing! Passage 2 may be familiar as the opening line of the *Star Trek* series. You might be tempted to mark this down as being less grammatical, because it contains a famous example of a split infinitive: 'to boldy go'. However, this kind of (most likely

mistaken) grammatical rule is not what cognitive psychologists are typically interested in: we do not wish to dictate what the best or most eloquent way of speaking is, we simply wish to understand how people really speak. In these terms splitting the infinitive is a perfectly acceptable and grammatical form of language. Sentence 3 is not grammatical by most definitions, but if a two-year-old said it to you, you would understand what they meant quite easily. Sentence 4 is in some ways the opposite of Sentence 3, in that it seems grammatical, yet meaningless. It was made famous by Noam Chomsky as an example of how syntax and semantics can be dissociable. Finally, Sentence 5 is clearly ungrammatical and pretty hard to extract any meaning from. After a while you may be able to work out that the sentence is a scrambled version of 'The words in this sentence have become rather jumbled'. It demonstrates just how important it is for us to have some mutually agreed conventions for word order, and this is precisely what syntax is!

Before embarking on a review of the models and data relevant to sentence processing, it is worth having a quick look at linguistic views of language structure. The constraints of our vocal and auditory systems dictate that words are uttered one by one in a serial fashion. However, according to many syntactic theories, this serial transmission obscures what is actually a hierarchical structure. Figure 6.8 illustrates the kind of syntactic structure that might be assigned to a simple sentence like 'The girl spotted the yacht'.



**Figure 6.8** A simple phrase structure tree

In this hierarchical analysis, each word is assigned a syntactic role in the sentence. The broken lines mark the links between each word and its role. These constituents are then grouped into phrases according to **phrase structure** rules, which are grammatical rules of English that indicate how phrases can combine. At the highest level the phrases combine to form a sentence. The analysis of sentences using **phrase structure grammar** offers a purely linguistic description, but we can see how it might apply to human language processing. On the perceptual side, we might think parsing should involve taking each word in a sentence, fleshing out its grammatical role and building a phrase structure tree that fits the sentence. This would not in itself recover the meaning of the sentence, but it would assist in this process, by facilitating the **thematic role assignment** for the sentence (i.e. identifying the fact that the girl is the do-er, identifying spotting as the activity she is doing, and that the yacht is what she is doing it to!).

Parsing is made more difficult by the fact that, as mentioned earlier, many words can be used in different grammatical roles. For example, the word *spotted* is used as a verb in Figure 6.8, but can be used as an adjective, as in a *spotted dress*. Equally, the noun *yacht* can be used as a verb, as in *to yacht around the world*. In the example sentence these alternative roles can be ruled out, because they would not form a grammatically coherent sentence in their alternative roles, but in many cases full sentences can be interpreted in different grammatically well-formed ways. Altmann (1998) gives the example of the sentence ‘Time flies like an arrow’, which has more than 50 grammatically permissible interpretations. For example, *time* appears to be acting as a noun but it could also be used as a verb, as in *to time an egg*. Similarly, *flies* is most obviously a verb, but it could act as a plural noun – someone could time some flies! Semantically, such an interpretation may make little sense, but it could still be grammatical – just like Chomsky’s famous example (Sentence 4 in Activity 6.3). In these cases we need to make use of more than just syntactic knowledge to resolve the ambiguity. The next section discusses how different models of parsing cope with ambiguities of this type. All models assume that we need to make use of multiple sources of information but they differ in terms of the priority of the different information types.

## 4.2 Models of parsing

We found in Section 2.1 that the language system makes good use of the short time it takes to say a word. As speech enters the perceptual system, the cohort of potential candidates is whittled down, ensuring minimal delay in retrieving a word’s meaning. One can ask the same question at the sentence level: does parsing assign a syntactic structure only at major syntactic boundaries (or even at the end of a sentence), or does it do so **incrementally**, refining the set of plausible syntactic structures every time a new word is recognized? It will not surprise you to learn that current models of sentence processing assume that parsing is incremental, and again this makes sense in terms of maximizing the availability of information for responding to the sentence. There are numerous demonstrations of incremental processing, employing a wide range of methods – an early example is the study of Tyler and Marslen-Wilson (1977). They made use of ambiguous phrases such as *landing planes*. With a preceding context such as ‘If you walk too near the runway, ...’ the natural interpretation of *landing* is as an adjective (e.g. ‘*landing planes* are dangerous’ would be a suitable continuation), whereas following ‘If you’ve been trained as a pilot, ...’ the interpretation is more likely to be as a verb (e.g. ‘*landing planes* is easy’). Tyler and Marslen-Wilson wanted to know whether listeners showed a contextual bias in their parsing of the ambiguous phrase. If parsing is delayed until a syntactic boundary is reached, then there should be no effect of preceding context on listeners’ expectations about whether the word following *landing planes* was either *is* or *are*. They gauged listeners’ expectations by presenting spoken fragments such as ‘If you’ve been trained as a pilot, *landing planes* ...’ to participants and asking them to name a visual target word (either *is* or *are* in this case). They found that the speed of a naming response depended on the preceding context of the ambiguous phrase. Appropriate continuations were named quickly, compared with inappropriate ones. This is incompatible with the ‘delayed parsing’ hypothesis, because such a model predicts no effect of appropriateness. Instead it fits in with the idea that a

plausible parse of a sentence is built up incrementally, and this influences expectations about upcoming words.

One of the most influential models of parsing, often known as the garden path model (Frazier, 1979), assumes that parsing is incremental, so each word is allocated a syntactic role as soon as it is perceived. In cases where more than one syntactic structure is compatible with the sentence so far, the parser makes a decision about which alternative to pursue based on syntactic information alone. The ‘garden path’ element comes in because the model predicts that there will be cases where the parse chosen at a point of ambiguity is incorrect (so the listener is ‘led down the garden path’). Later in the sentence this incorrect selection will become clear, causing some backtracking as an alternative interpretation is attempted. The idea of pursuing some hypothesis and then reaching a dead-end requiring re-analysis fits in with people’s intuitions about how they interpret some sentences. A famous example of ‘garden pathing’ is the sentence: ‘The horse raced past the barn fell’ (Bever, 1970). As you read this sentence, you may have had problems integrating the final word. Some people think that maybe there is an ‘and’ missing between *barn* and *fell*, or that there is a comma missing between *past* and *the*. But there is an alternative interpretation, which is a reduced version of ‘The horse that was raced past the barn fell’. According to the garden path model, this alternative is not chosen when the word *raced* is first perceived, leading to trouble with interpretation later in the sentence.

The garden path model makes use of a set of guiding principles that specify which parse should be selected in the case of syntactic ambiguities, and these principles involve only syntactic information. The details of these principles are not essential – it is more important to keep in mind that the garden path model assumes a *serial* parser that maintains only one potential parse of a sentence at a time, and has an *autonomous* component, in that the initial evaluation of a word’s role in a sentence is based only on syntactic factors. In direct contrast, constraint-based models (e.g. MacDonald *et al.*, 1994) assume that parsing is *parallel* and *interactive*. So rather than maintaining a single syntactic analysis, these models allow more than one potential parse of a sentence to be evaluated at the same time (just as the cohort model of word recognition evaluates numerous candidates for word identification). Constraint-based models are thought of as interactive because they eliminate the autonomous stage of parsing assumed by the garden path model. Instead, other factors, such as frequency and semantic plausibility can influence parsing immediately.

MacDonald *et al.*’s model also increases the involvement of the lexicon in the parsing process, by assuming that some information about how a word can combine with other words is stored in the lexicon. By this kind of account, parsing becomes a bit like fitting the pieces of a jigsaw together. Each piece contains information about a word, including the kinds of syntactic context the word could fit into, and parsing involves fitting all the pieces together so that the words form a coherent sentence.

The two models described here are by no means the only models of parsing that researchers currently consider, but they do mark out the kinds of properties that generate debate in this area, and they highlight the kinds of questions that we need to investigate through experimentation. First and foremost among these, we need to try to address the question of whether parsing is autonomous, or whether it makes use of non-syntactic sources of information stored in the lexicon.

### 4.3 Is parsing autonomous?

As we have seen, the garden path model makes the strong prediction that the initial syntactic analysis of a word is unaffected by factors such as the meaning of the preceding context, or the meaning of the words. In essence, the model puts all aspects of semantics aside until a word has been assigned a syntactic role. Initial data on the resolution of syntactic ambiguity showed garden path effects fully consistent with the autonomous approach of Frazier's model. In addition, some experiments designed specifically to look for semantic influences on syntactic ambiguity resolution found none. Ferreira and Clifton (1986) investigated how readers interpret verbs in phrases such as 'The defendant examined ...'. Before you read on, think about how you might continue this sentence fragment. There are two common roles that 'examined' can play in this context. It could simply be the main verb of the sentence, as in 'The defendant examined his hands'. But it could also form part of what is known as a **reduced relative clause**. A relative clause might be 'The defendant that was examined by the lawyer ...', and the reduced form would simply be the same but with 'that was' eliminated. The garden path model states that the preferred structure when *examined* is encountered is the more straightforward main verb interpretation. So if the sentence continuation is in fact a reduced relative structure, the Frazier model predicts a garden path effect when the true structure of the sentence becomes clear. So when a reader encounters 'The defendant examined by the lawyer turned out to be unreliable' they should show evidence of processing difficulty. This is exactly what Ferreira and Clifton found, using the eye-tracking methodology – people tended to fixate on the region just after the ambiguity, suggesting that they were having trouble incorporating the new information into their initially selected parse of the sentence.

The critical question here was whether the meaning of the word preceding the ambiguous verb could affect the garden path effect. So Ferreira and Clifton compared sentences like the one above to sentences like 'The evidence examined by the lawyer turned out to be unreliable'. In this case *evidence* is inanimate, which reduces the plausibility of the main verb interpretation (i.e. it seems unlikely that the evidence would examine anything). Despite the semantic bias towards the alternative reading, the garden path effect remained (i.e. fixation times remained long). On the surface, this seems like sound support for the autonomy assumed by the garden path model.

However, Trueswell *et al.* (1994) noticed that some of the contexts used by Ferreira and Clifton were less constraining than the example above. It is difficult to imagine a situation in which evidence could examine something, but Trueswell *et al.* argued that this was not the case for about half the materials used in the original experiment (e.g. 'the car towed ...' where *car* is inanimate, but still quite a plausible candidate for something that tows). They ran another eye-tracking experiment using a similar design, but with more constraining semantic contexts, and found that these contexts could lessen or even eliminate the garden path effect. The results of this and other similar studies are important because they show that, in some circumstances, the parsing system can be strongly affected by the semantic plausibility of the various parses of the system. The garden path model could perhaps be saved if the autonomous parsing component is assumed to last only a short time, and that other factors come into place soon afterwards, but this greatly weakens the predictive

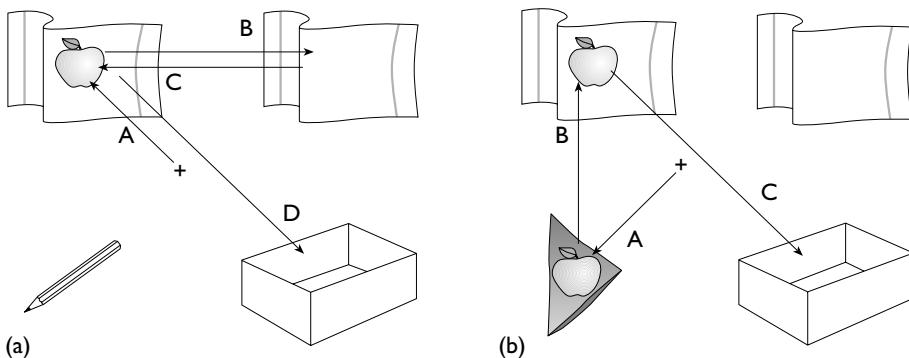
power of the model, because it becomes harder to distinguish from models which allow semantic factors to play a stronger role. But it is worth remembering that Frazier's syntactic constraints are not rendered immaterial by the finding that parsing is influenced by semantic plausibility. Instead, syntactic constraints appear to operate in *combination* with other constraints, with the ultimate goal being to weigh up the likelihood of different parses of a sentence in cases of ambiguity.

#### 4.4 Constraints on parsing

It seems that the parsing system can be influenced by quite a number of different factors when it encounters an ambiguity. When a sentence is spoken, there is often useful information in the rhythm of the sentence. Think about how you might say the sentence 'Jane hit the man with the hammer' in the cases where (a) the man has a hammer or (b) Jane has a hammer. One way to distinguish between these two possibilities is by changing your speech rate mid-sentence, so that different sets of words are grouped together. Of course these changes will be exaggerated when the speaker is aware of the potential ambiguity, but even in normal speech, the speaker can reduce ambiguity with changes in pitch and timing, and the listener can make use of this information (Warren, 1996). At a very different level, information about how often words are used in different syntactic structures can also influence the parsing process. This factor can be seen at work in the earlier example from Bever (1970), 'The horse raced past the barn fell'. One of the reasons this sentence causes so many problems is because the verb race is rarely used as a **past participle** (i.e. as in 'the horse that was raced ...'). Not all verbs have this strong bias, so for example *released* has the opposite bias – it is more likely to be used as a past participle (e.g. 'The hostage was released') than as a past tense of a main verb (e.g. 'The government released a press statement'). Trueswell (1996) showed that this lexical frequency factor also influenced the way in which sentences are parsed. People seem to be able to keep track of the ways in which words are used in different sentences, and apply this knowledge in cases of ambiguity.

Perhaps the most striking example of a contextual influence on syntactic processing is based on the use of visual information. Tanenhaus *et al.* (1995) wanted to know whether the visual context of a sentence would affect the interpretation of syntactic ambiguities. In order to do this, they sat participants at a table on which some objects like apples and towels were placed, and gave them instructions to move the objects such as 'Put the apple on the towel in the box'. The participants wore head-mounted eye trackers so that the experimenters could monitor eye movements as the sentences were heard (see Box 6.2 in Section 2.2.3). The sentences had a temporary syntactic ambiguity, which in the case of the example here involves the phrase 'on the towel'. We know from studies like Ferreira and Clifton (1986) that when people hear 'Put the apple on the towel ...' they tend to interpret 'on the towel' as the desired destination of the apple. But the continuation '... in the box' should force a reassessment of the sentence (i.e. the sentence is a reduced form of 'Put the apple that's on the towel in the box'). We have seen that various sentential factors such as semantic plausibility can reduce or eliminate this garden path effect, but what about external, environmental context? Tanenhaus *et al.* (1995) gave the participants instructions in two types of external context (see Figure 6.9). In one case (see Figure 6.9(a)), there was an apple on a towel, another towel, and a box. This context

supports the initial interpretation of ‘on the towel’ as referring to the destination, so people tended to look at the apple and then the empty towel, and only looked at the true intended destination once the disambiguating speech (‘in the box’) was heard. However, when the scene also included a second apple on a napkin (see Figure 6.9(b)) participants’ eye movements were quite different. Now when they heard ‘on the towel’, they rarely looked at the empty towel, because they interpreted ‘on the towel’ as distinguishing information between the two apples (one was on a towel and one on a napkin). In other words, the environmental situation provided a source of information that could eliminate the garden path effect.



**Figure 6.9** Two visual contexts from Tanenhaus et al. (1995) showing the typical sequence of eye movements in response to the ambiguous instruction ‘Put the apple on the towel in the box’. Eye fixations began at the central cross, and continued in the sequence indicated by the capital letters

Source: based on Tanenhaus et al., 1995, Figures 1 and 2

## Summary of Section 4

---

- Understanding a sentence requires a parsing process in which each word is assigned a grammatical role.
- The garden path model assumes that the parser operates autonomously, without any influence of nonsyntactic factors.
- Recent studies of syntactic ambiguity resolution suggest that a variety of different constraints can influence parsing, including even the environment of the listener.

## 5 Conclusion

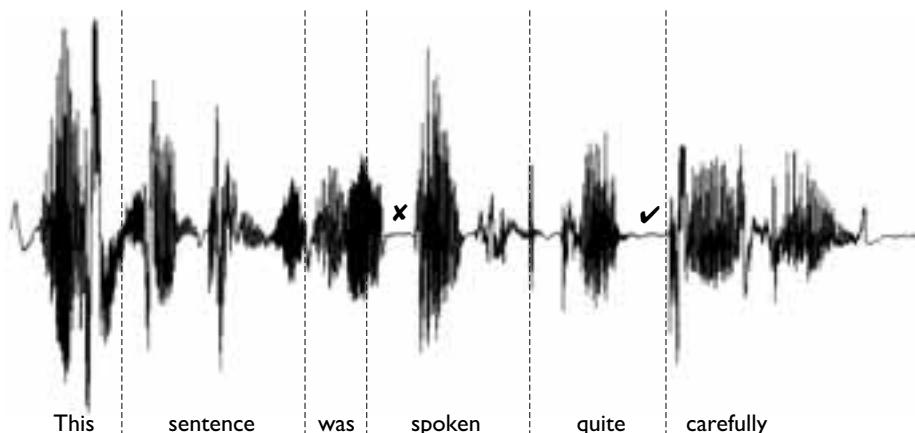
This chapter has provided a brief account of some of the main components of the language system, particularly with reference to recognizing words and understanding sentences. We have seen that many of the processes involved can be modelled in terms of competition between multiple candidates, implying that the language system is busy evaluating countless hypotheses about an utterance at numerous levels, at any moment in time. Thankfully we remain blissfully unaware

of these operations, with only a pretty terse ‘executive summary’ of the process available to conscious awareness.

Another recurring theme has involved the extent to which components of the system operate independently of each other. There is a long way to go in this debate, but the current state of play seems to be one in which there is a surprising level of linkage between subsystems. So reading a word engages processes and representations related to speech perception, and the way in which you process a spoken sentence can be influenced by the real world context in which you hear it. This interconnectedness may well reflect two aspects of language processing: the complexity of language, and the speed with which we need to communicate. In terms of language development, it makes a lot of sense to re-use existing mechanisms when we are trying to add a new mechanism such as the mechanism for reading. In terms of adult language processing, it makes sense to call on as much useful information as possible to minimize the time it takes to comprehend a sentence.

### Answer to Activity 6.1

The approximate word boundary positions are marked in Figure 6.10, along with the words themselves. Some gaps in the speech (low amplitude signal) are aligned with word boundaries (e.g. between *quite* and *carefully*, marked ✓), whereas others are not (e.g. within *spoken*, marked ✗). In general, short periods of silence are poor indicators of word boundaries, meaning that we have to find better ways to segment speech.



**Figure 6.10** Typical word boundaries in a fluent sentence

### Further reading

Altmann, G.T.M. (2001) ‘The mechanics of language: psycholinguistics in review’, *British Journal of Psychology*, vol.92, pp.129–70.

Harley, T. (2001) *The Psychology of Language: From Data to Theory*, Hove, Psychology Press.

## References

- Altmann, G.T.M. (1998) 'Ambiguity in sentence processing', *Trends in Cognitive Sciences*, vol.2, no.4, pp.146–52.
- Berko, J. (1958) 'The child's learning of English morphology', *Word*, vol.14, no.2, pp.150–77.
- Bever, T.G. (1970) 'The cognitive basis for linguistic structures', in Hayes, J. (ed.) *Cognition and the Development of Language*, New York, Wiley.
- Collins, A.M. and Loftus, E.F. (1975) 'A spreading-activation theory of semantic processing', *Psychological Review*, vol.82, no.6, pp.407–28.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. and Ziegler, J. (2001) 'DRC: a dual route cascaded model of visual word recognition and reading aloud', *Psychological Review*, vol.108, no.1, pp.204–56.
- Cutler, A., and Carter, D.M. (1987) 'The predominance of strong initial syllables in the English vocabulary', *Computer Speech and Language*, vol.2, no.2, pp.133–42.
- Cutler, A. and Norris, D. (1988) 'The role of strong syllables in segmentation for lexical access', *Journal of Experimental Psychology: Human Perception and Performance*, vol.14, no.1, pp.113–21.
- Cutler, A. and Otake, T. (2002) 'Rhythmic categories in spoken-word recognition', *Journal of Memory and Language*, vol.46, no.2, pp.296–322.
- Ferreira, F. and Clifton, C. (1986) 'The independence of syntactic processing', *Journal of Memory and Language*, vol.25, no.3, pp.348–68.
- Frazier, L. (1979) 'On comprehending sentences: syntactic parsing strategies', *Bloomington*, IN, Indiana University Linguistics Club.
- Frost, R. (1998) 'Toward a strong phonological theory of visual word recognition: true issues and false trails', *Psychological Bulletin*, vol.123, no.1, pp.71–99.
- Gaskell, M.G. and Marslen-Wilson, W.D. (2002) 'Representation and competition in the perception of spoken words', *Cognitive Psychology*, vol.45, no.2, pp.220–66.
- Glushko, R.J. (1979) 'The organization and activation of orthographic knowledge in reading aloud', *Journal of Experimental Psychology: Human Perception and Performance*, vol.5, no.4, pp.674–91.
- Grainger, J. and Jacobs, A.M. (1994) 'A dual read-out model of word context effects in letter perception – further investigations of the word superiority effect', *Journal of Experimental Psychology: Human Perception and Performance*, vol.20, no.6, pp.1158–76.
- Hauser, M.D., Newport, E.L. and Aslin, R.N. (2001) 'Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins', *Cognition*, vol.78, no.3, B53–B64.
- Kellenbach, M.L., Wijers, A.A. and Mulder, G. (2000) 'Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming', *Cognitive Brain Research*, vol.10, nos1–2, pp.67–75.

- Kutas, M. and Hillyard, S.A. (1980) 'Reading senseless sentences: brain potentials reflect semantic incongruity', *Science*, vol.207, no.4427, pp.203–5.
- Lucas, M. (1999) 'Context effects in lexical access: a meta-analysis', *Memory and Cognition*, vol.27, no.3, pp.385–98.
- Lucas, M. (2000) 'Semantic priming without association: a meta-analytic review', *Psychonomic Bulletin and Review*, vol.7, no.4, pp.618–30.
- MacDonald, M.C., Pearlmuter, N.J. and Seidenberg, M.S. (1994) 'Lexical nature of syntactic ambiguity resolution', *Psychological Review*, vol.101, no.4, pp.676–703.
- Marslen-Wilson, W.D. (1987) 'Functional parallelism in spoken word recognition', *Cognition*, vol.25, no.1, pp.71–102.
- Marslen-Wilson, W.D. and Welsh, A. (1978) 'Processing interactions and lexical access during word recognition in continuous speech', *Cognitive Psychology*, vol.10, no.1, pp.29–63.
- Marslen-Wilson, W.D., Tyler, L.K., Waksler, R. and Older, L. (1994) 'Morphology and meaning in the English mental lexicon', *Psychological Review*, vol.101, no.1, pp.3–33.
- Masson, M.E.J. (1995) 'A distributed memory model of semantic priming', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, no.1, pp.3–23.
- McClelland, J.L. and Elman, J.L. (1986) 'The TRACE model of speech perception', *Cognitive Psychology*, vol.18, no.1, pp.1–86.
- McClelland, J.L. and Rumelhart, D.E. (1981) 'An interactive activation model of context effects in letter perception. Part 1: an account of basic findings', *Psychological Review*, vol.88, no.3, pp.375–407.
- McConkie, G.W. and Rayner, K. (1975) 'The span of the effective stimulus during a fixation in reading', *Perception and Psychophysics*, vol.17, no.6, pp.578–86.
- McQueen, J.M., Norris, D. and Cutler, A. (1994) 'Competition in spoken word recognition – spotting words in other words', *Journal of Experimental Psychology: Learning Memory and Cognition*, vol.20, no.3, pp.621–38.
- Nelson, D.L., McEvoy, C.L. and Schreiber, T.A. (1998) *The University of South Florida Word Association, Rhyme, and Word Fragment Norm* [online]. Available from <http://w3.usf.edu/FreeAssociation> (accessed 19 January 2004).
- Norris, D., McQueen, J.M. and Cutler, A. (2000) 'Merging information in speech recognition: feedback is never necessary', *Behavioral and Brain Sciences*, vol.23, no.3, pp.299–370.
- O'Regan, J.K. and Jacobs, A.M. (1992) 'Optimal viewing position effect in word recognition – a challenge to current theory', *Journal of Experimental Psychology: Human Perception and Performance*, vol.18, no.1, pp.185–97.
- Rayner, K. and Duffy, S.A. (1986) 'Lexical complexity and fixation times in reading – effects of word-frequency, verb complexity, and lexical ambiguity', *Memory and Cognition*, vol.14, no.3, pp.191–201.

- Rodd, J., Gaskell, G. and Marslen-Wilson, W. (2002) 'Making sense of semantic ambiguity: semantic competition in lexical access', *Journal of Memory and Language*, vol.46, no.2, pp.245–66.
- Saffran, J.R., Aslin, R.N. and Newport, E.L. (1996) 'Statistical learning by 8-month-old infants', *Science*, vol.274, no.5294, pp.1926–8.
- Saffran, J.R., Johnson, E.K., Aslin, R.N. and Newport, E.L. (1999) 'Statistical learning of tone sequences by human infants and adults', *Cognition*, vol.70, no.1, pp.27–52.
- Seidenberg, M.S., and McClelland, J.L. (1989) 'A distributed, developmental model of word recognition and naming', *Psychological Review*, vol.96, no.3, pp.523–68.
- Shillcock, R., Ellison, T. M. and Monaghan, P. (2000) 'Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model', *Psychological Review*, vol.107, no.4, pp.824–51.
- Starr, M., and Rayner, K. (2001) 'Eye movements during reading: some current controversies', *Trends in Cognitive Sciences*, vol.5, no.4, pp.156–63.
- Swinney, D.A. (1979) 'Lexical access during sentence comprehension: (re)consideration of context effects', *Journal of Verbal Learning and Verbal Behavior*, vol.18, no.5, pp.645–59.
- Tabossi, P. and Zardon, F. (1993) 'Processing ambiguous words in context', *Journal of Memory and Language*, vol.32, no.3, pp.359–72.
- Taft, M. and Forster, K.I. (1975) 'Lexical storage and retrieval of prefixed words', *Journal of Verbal Learning and Verbal Behavior*, vol.14, no.6, pp.638–47.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. and Sedivy, J.C. (1995) 'Integration of visual and linguistic information in spoken language comprehension', *Science*, vol.268, no.5217, pp.1632–4.
- Trueswell, J.C. (1996) 'The role of lexical frequency in syntactic ambiguity resolution', *Journal of Memory and Language*, vol.35, no.4, pp.566–85.
- Trueswell, J.C., Tanenhaus, M.K. and Garnsey, S.M. (1994) 'Semantic influences on parsing – use of thematic role information in syntactic ambiguity resolution', *Journal of Memory and Language*, vol.33, no.3, pp.285–318.
- Tyler, L.K. and Marslen-Wilson, W.D. (1977) 'The on-line effects of semantic context on syntactic processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, no.6, pp.683–92.
- Van Orden, G.C. (1987) 'A rows is a rose – spelling, sound, and reading', *Memory and Cognition*, vol.15, no.3, pp.181–98.
- Van Petten, C. and Kutas, M. (1987) 'Ambiguous words in context – an event-related potential analysis of the time course of meaning activation', *Journal of Memory and Language*, vol.26, no.2, pp.188–208.
- Warren, P. (1996) 'Prosody and parsing: an introduction', *Language and Cognitive Processes*, vol.11, nos 1–2, pp.1–16.

# Language in action

## Chapter 7

Simon Garrod and Anthony J. Sanford

---

### 1 Introduction

In the previous chapter we saw how language comprehension can be viewed as a process of *constructing an interpretation*. Linked language processing subsystems are involved in matching candidate interpretations against the input, until a plausible meaning is selected. In this chapter, we go beyond this view of language processing to look at language in action in everyday settings and examine how models of language comprehension and language production need to reflect the different circumstances under which language is used.

Language is used primarily for verbal communication and more often through speech rather than through writing. From the moment you get up in the morning until you finally fall asleep at night you will very often be speaking to someone or other, and sometimes perhaps almost every minute of the day. Furthermore, when you do this you are nearly always engaging in a dialogue, in which one or more people interact directly with you and each other. Of course, you can also use written language to communicate. Usually this will be through monologue, in which there is no interaction between the writer and the reader, as when you write an essay or read a newspaper article. But increasingly even written communication is becoming interactive. Consider, for instance, exchanging text messages with a friend or taking part in a ‘chat room’ conversation on your computer. Whether it is through speech or writing, communication is perhaps the most important social, cultural and cognitive activity that we engage in, so understanding how we use language to communicate is central to the study of human cognition.

Although it can easily be argued that spoken language, and in particular dialogue, is the more basic form of language use, we begin this chapter with a look at research on written language, and only then proceed to dialogue. This sequence reflects the history of experimental research on the psychology of language. Much more research has been done on written language comprehension than on dialogue, and the findings of this research illustrate important features of language in action that take us beyond the simple construction metaphor discussed in Chapter 6.

One way in which written language comprehension takes us beyond the processes discussed in Chapter 6 is that it requires the integration of information across different sentences in a text. As we shall see below, a key aspect of this integration is that it depends upon access to non-linguistic background knowledge. To this extent, language processing in the broader sense is less ‘encapsulated’ than it seemed in the previous chapter. The precise interpretation of any real piece of language calls upon a wide range of different sources of information, including our knowledge of the situation under discussion.

We begin by defining what is meant by text and then consider what this means for text processing.

## 2 Written language and discourse

Traditionally linguists have identified two characteristics that differentiate a text from just a collection of isolated sentences. The first is what they call **cohesion**; the second is **coherence** (Halliday and Hasan, 1976; Brown and Yule, 1983).

Texts are cohesive to the extent that they contain many expressions whose interpretation depends in some way on interpretations of prior expressions in the text and these co-interpretations serve to link the sentences together. One major source of cohesion comes from **anaphora** (repeated reference). For instance, the sentences in the pair below are cohesive because the pronouns *she* and *it* in 2 take their interpretation from the noun-phrases *Susan* and *some money* in 1:

- 1 Bill wanted to lend Susan some money.
- 2 She was hard up and really needed it.

Furthermore, the cohesive link contributes to the fact that sentences 1 and 2 constitute a piece of text. But cohesion is not all there is to bind sentences together into a text. For instance, consider the following variant of sentences 1 and 2:

- 1' Bill wanted to lend Susan some money.
- 2' It is not nice to have close friends who are really hard up.

Here there are no cohesive anaphoric links between the sentences yet we still have an acceptable text. What is important in this case is that the two sentences can be related into a coherent whole through inference. The reader will take it that the unpleasantness of having friends who are hard up is the reason why Bill wants to lend Susan some money and, by implication, that Susan is Bill's friend. So a text's coherence comes from establishing the logical and psychological consistency between the events and mental states portrayed (with respect to the intentions of the characters in the text).

Cohesion and coherence are not independent. Even in texts such as 1'-2' above there is a kind of cohesive bond set up because it is assumed that 'Susan' must be an instance of one of 'Bill's close friends'. In fact, it will often be the case that the interpretation of cohesion markers, such as pronouns, depends upon establishing coherence, and vice versa. Consider, for example, the following further variant of sentences 1 and 2:

- 1'' Bill wanted to lend his friend some money.
  - 2'' He was hard up and really needed it.
- and
- 2''' However, he was hard up and couldn't afford to.

The same pronoun *he* in almost identical clauses (2'' and 2''') takes on different referential interpretations depending upon the different coherence relations between the two sentences. At the same time, the form of coherence relation differs depending on the assignment of the pronoun. For instance, while his (the friend's) being hard up in 2'' is taken as a reason for Bill's wanting to lend money, his (Bill's) being hard up in 2''' is taken as an obstacle to Bill's wanting to lend the money.

Therefore, collections of sentences become texts through the links that bind them together into a coherent structure. Some of the links are signalled explicitly through cohesion markers, such as pronouns or sentence connectives like *but*, *therefore*, *however*, whereas other links depend upon inferring the logical or psychological relationships between the events portrayed. Besides reference, there are many other sources of linkage. For instance, in narrative text there have to be temporal links that order the events in the story. In simple cases these are signalled with explicit temporal expressions as in the following short passage:

- 3 Yesterday, Mary visited ( $e_1$ ) her grandmother. Later, she stopped ( $e_2$ ) at a shop to buy some flowers. ( $e_1$  and  $e_2$  denote events.)

Here the events are explicitly ordered through the temporal cohesion device *later*. So event  $e_1$  precedes event  $e_2$ . But again, ordering often comes from establishing a coherent chain of events. For example, in the following variant the ‘visiting’ and the ‘stopping at the shop’ are interpreted as occurring in the opposite order:

- 4 Yesterday, Mary visited ( $e_2$ ) her grandmother. She stopped ( $e_1$ ) at a shop to buy some flowers. She then went and presented them to her as a gift for her eightieth birthday.

So temporal cohesion, like referential cohesion, often depends upon the coherence of the passage as a whole.

Examples like those above where there is no explicit marker to indicate how the sentences relate to each other suggest that the coherence of a discourse is in the mind of the reader. In all of the examples above the reader uses general knowledge to make a coherent connection. Brown and Yule (1983) contrast the discourse-as-product view (the coherence is in the text alone) with discourse-as-process, where the coherence comes from mental processes called upon to interpret the text. The second view leads us naturally to psychological investigations of interpretation, and to an examination of how the mind adds to what is in the text.

## 2.1 Processes underlying text interpretation

As we have seen from Chapter 6, much research on language comprehension is concerned with how sub-processes operate in real time, and establishing when each sub-operation occurs that eventually leads to a coherent understanding of the text. Here we meet a number of issues, each of which concerns the establishment of coherence, and extracting the meaning of discourse beyond just the meaning of the words it contains.

### 2.1.1 Anaphora resolution

As we have already seen, anaphoric reference is crucial to text cohesion. So it is not surprising that how we resolve anaphors during comprehension is one of the most studied components of text comprehension (Garnham, 2001). When reading a text, it is important to keep track of who is doing what when. For instance, given the passage in Table 7.1, we need to know what *John* did and what *William* did.

**Table 7.1** A sample of coherent text (we shall use this example to discuss a number of processing issues throughout the chapter)

---

John went to the shops with Mary. She went off to buy clothes, and he went to the bank. With the money he had, he was going to buy some new CDs. On the way to the shop, he bumped into William. He found out William hadn't had any lunch that day. John lent William some money because he was hard up.

---

Right until the last sentence, every time we encounter *he*, we interpret it as standing for *John* (it is an anaphor for *John*). He does not stand for *Mary*, because *Mary* is female, and *he* specifies a singular and male antecedent. There is no ambiguity, and we might suppose that the language processor checks to see what *he* stands for as soon as *he* is encountered. *John* is the main character in the story, being mentioned often, and so even when a new male individual turns up there is a preference to equate *he* with *John*.

In the last sentence of the text in Table 7.1, something different happens. Here *he* stands for *William*. The new pronoun assignment works because the reader can use his or her general knowledge to disambiguate the pronoun. As we saw earlier the reader can assume that people who have money are in a position to lend, and people who are hard up need money. So, putting these facts together with what the text says enables the processor to resolve *he* as referring to *William*.

In this hypothetical analysis, we can see several potential sources of information that the comprehension system might use to resolve pronoun-based (pronominal) anaphora:

- Gender cues in the pronouns. Apart from gender marking (*he* vs. *she*), we have animacy marking (*he*, *she* vs. *it*), reflexivity (*he* vs. *himself*, *she* vs. *herself*).
- Main character vs. secondary character (or continued reference to one character).
- General knowledge: for example, people with X can give X, people without can't. People without X might want X, and so on. These properties form a very large set of general knowledge beliefs.

These are just some of the cues that are known to support anaphora resolution. A full discussion of anaphora from a psychological standpoint is given in Garnham (2001).

### 2.1.2 When word meaning is used

The text in Table 7.1 illustrates other problems that psycholinguists have worked on intensively. One is the question of when it is that word meaning enters into the comprehension process. As you saw in Chapter 6, a useful technique for studying when meaning may be accessed is to track a person's eye movements during reading. To find out when word meaning is accessed, a word that is not appropriate, or anomalous, is inserted into the text to see if it disturbs the pattern of eye movements as soon as it is encountered. If it does (causing longer initial fixations, or causing an increase in regressive eye movements), then it can be concluded that the word's meaning is being used *at that point*.

- Traxler and Pickering (1996) compared how people read materials like 5 and 6:
- 5 That's the pistol with which the man shot the gangster yesterday afternoon.
  - 6 That's the garage with which the man shot the gangster yesterday afternoon.

The word *shot* fits sentence 5 in meaning, but not sentence 6. Participants read sentences of this type while having their eye movements monitored. Traxler and Pickering found that the very first fixations on the word *shot* were longer in the implausible cases, like 6, than in the plausible cases, like 5. So, the meaning of *shot* must have been accessed, and incorporated into the meaning of the sentences as a whole as soon as the word was fixated. Results such as these are consistent with what is called **incremental interpretation**, a view of discourse comprehension that says each word is interpreted and incorporated into the meaning of the sentence as soon as it is encountered. (You met this in Chapter 6, Section 4.2 when parsing was discussed.)

### 2.1.3 Non-literal meaning

The immediacy of processing observed in the studies above is consistent with what one might term the ‘standard view’ of text understanding. In this view, as words are encountered, their meanings are retrieved from long-term (semantic) memory. As the sentence unfolds, the syntactic structure is derived, and the meanings of the words are then combined to give a sentence meaning. However, there are several problems with this view. One is the problem of non-literal meaning. Consider the sentence, ‘John asked the man if he could tell him the time’. Although ‘Can you tell me the time?’ is a literal question (to which the answer is ‘yes’ or ‘no’), the interpretation given is as a request (‘Tell me what the time is, please’). This is an example of what is known as an *indirect* speech act. How do people get the correct interpretation? The standard account has the following steps (Glucksberg and Keysar, 1990):

- Derive a literal interpretation.
- Assess that interpretation against the context of the utterance.
- If and only if literal meaning is a poor fit, derive a non-literal interpretation.

This suggests that indirect speech acts should take longer to process than direct speech acts (e.g. questions that are in fact questions and not indirect requests). Certainly for some cases, the model is wrong, because the indirect cases are processed just as quickly as the direct cases (e.g. Gibbs, 1983). Now look at the following sentence from Table 7.1: ‘On the way to the shop, he bumped into William ...’ This is interpreted as John meeting William, probably unexpectedly, and not actually colliding with him. It requires nonliteral interpretation. Now according to the model above, a statement will only get a nonliteral interpretation if it is needed. Glucksberg *et al.* (1982) showed that nonliteral interpretations may be given when they are *not* needed, suggesting that they are derived automatically. They asked participants to decide whether statements were literally true; for instance, ‘Some desks are junkyards’. What they showed was that the time to answer ‘no’ was longer in such cases than in the case of literally false statements with no metaphoric interpretation, such as ‘Some desks are roads’. They suggest that this is because the nonliteral meaning is highly available from memory, and so it intrudes in the ‘true’ judgement, leading to longer times to decide that it is not literally true.

It might be thought that such cases are rare, but they are not. In particular, Lakoff (1987) has provided numerous examples of metaphors used in everyday language. Here are a few:

- 7 When John heard about his wife, he exploded.
- 8 When it came to chemistry, Fred was a little rusty.
- 9 It is morally right to fight poverty.

## ACTIVITY 7.1

---

If you think about the literal meaning of the words in these sentences, you will see quickly that they contain metaphors. Try to think of your own examples of such things. Do you think they provide a challenge for a view that says sentence meaning is based on the literal meanings of words?

---

### 2.1.4 Inferences

We have seen that general knowledge is needed to understand texts. The processes that give rise to coherence are known as **inference-making**. In psycholinguistics, a distinction has been drawn between **necessary inferences** and **elaborative inferences**. Here is a case where a necessary inference allows interpretation:

- 10 Mary got some picnic things out of the trunk of the car. The beer was warm.
- 11 Mary got some beer out of the trunk of the car. The beer was warm.

The inference in 10 is that the beer is part of the picnic things, and so came out of the trunk of the car. Haviland and Clark (1974) were the first to show that such inferences take time. Using self-paced reading, they showed that reading times for the *second* sentence in examples such as 10, where beer hadn't been mentioned before, were longer by some 100 msec than in examples such as 11, where it had. The extra time is the time to form the inference *beer is part of the picnic things*, which forms a link, or bridge, and so is sometimes called a **bridging inference**. Unless the inference had been drawn, there would be no way in which the two sentences could be sensibly connected – it would not be a coherent discourse. That is why such inferences are called necessary.

Elaborative inferences refer to inferences that are not strictly necessary. Consider 12:

- 12 Unable to control his rage, the angry husband threw the valuable porcelain vase at the wall.

Did you make the inference that the vase broke? It is certainly a plausible inference, as are several others, like 'He was having a row with his wife'. Such inferences are not necessary for understanding the sentence. In fact, they are **defeasible** (can be cancelled). For instance, if you read 'He missed and the vase landed on the sofa', you would have to cancel your inference that the vase broke. Research into whether elaborative inferences are made has been carried out by using a variety of priming techniques, such as showing the word *broke* after 12. For example, if the word *broke* were to be primed in a lexical decision task, then that would suggest that the inference had been made. The evidence suggests that elaborative inferences are not made regularly (McKoon and Ratcliff, 1992), but it

remains to be shown exactly what are the conditions under which they are and are not made.

### 2.1.5 Relating language to knowledge

Almost all of the processes discussed above show how world knowledge is needed to interpret the meaning of what is being said in a text. This leads to a variety of questions concerning the text–knowledge interface. One very important concept is that understanding depends upon the reader setting up a mental model of what the text is depicting. Take a look at examples 13 and 14:

- 13 Harry put the wallpaper on the table. Then he sat his cup on it.
- 14 Harry put the wallpaper on the wall. Then he sat his cup on it.

Sentence 14 sounds odd because when wallpaper is on the wall, it is in a vertical plane, and would not support a cup. If you noticed this, which you almost certainly will have, then this means you produced a mental model of what the first sentence meant in relation to the real world, and when you integrated the second sentence, there was a problem. Approaches to language understanding all include some reference to how world knowledge is involved in interpretation (e.g. Kintsch, 1988; Sanford and Garrod, 1998). Here we want to emphasize that it is very important for readers to have the right mental model if they are to understand the discourse. Sanford and Garrod (1981, 1998) believe that much of our knowledge is organized in situation-specific packages. For instance, if one reads about buying something in a shop, the important aspects of what this entails become available as part of one's mental representation of what the text is about. This means, for instance, that the writer can refer to things that have not been explicitly mentioned before, as in 15:

- 15 The court case was going badly for the defendant. He could see that the judge had no time for him.

Despite there being no previous mention of a judge, encountering the phrase *the judge* causes no difficulty, and the processing time is no longer than if the judge had been mentioned explicitly (Sanford and Garrod, 1998 give further details).

These examples show how language accesses information in our memory that represents situations and settings. For instance, our memory contains information about what or who to expect to be present at a court case. Precisely *how* such situational information is represented has been a considerable area of inquiry. As we shall see below, one intriguing idea is that understanding relies on representations that are literally of how our bodies interact with the world. This is quickly becoming a key issue in how understanding works, and what meaning might be.

### 2.1.6 Knowledge, meaning and embodiment

In the traditional view of language processing, concepts are treated in an abstract way. Indeed, the meanings of words are commonly thought of as being represented as mental lists or networks of attributes (see Section 3, Chapter 6). According to this view, language conveys meaning by using abstract words, combined by syntactic rules (e.g. Fodor, 2000; Kintsch, 1988; Pinker, 1994). However, an alternative view has been emerging, in which meaning is rooted in perception and bodily action –

literally, in how we interact with the world. One motivation behind this embodiment view of meaning comes from what has been called the **symbol grounding problem** (Harnad, 1990).

One version of the argument is as follows, adapted from Glenberg and Robertson (2000). Suppose you travel to China and at the airport see a sign written in Chinese. All you have is a dictionary; nobody speaks to you. The first part of the sign can be found in the dictionary, so you look at the entry to see what it means. Of course, all you find is more Chinese script. You can repeat the process with the first part of the script, but that only continues the problem. No matter how many times you look the scripts up, you can never recover the original meaning. The dictionary does not contain the meaning of the expression.

Therefore, according to Harnad (1990) and Searle (1980), symbols can only have meaning by being related to things in the world, and not to other symbols and words. Consider the words *left* and *right*. Definitions for these in dictionaries make interesting reading. One entry for *left*, for instance, is that it is the opposite of right. Without some means of interpreting that statement, it simply doesn't make sense. Typically in dictionaries, there is reference to the outside world. So, *left* is defined as 'that side in which a person has normally the weaker and less skilled hand', and *right* as 'that side in which a person has normally the stronger and more skilled hand'. Unless one knows about people, and which side (regardless of name) is normally stronger, the definitions are vacuous. So, the meanings of words have to be grounded in the world. Consider another case, the verb *trudge*.

## ACTIVITY 7.2

---

Try to write down a definition of *trudge*.

Now try to write down a definition of *waltz*.

### COMMENT

You probably found that a verbal redescription of *trudge* was virtually impossible to produce. Even if you could produce something, you probably felt that the redescription was inadequate as a definition. Maybe you would have found it easier to show what *trudge* means by physically trudging. This is because *trudge* defines a set of motion attributes that are embodied in (human) movement. This too is obvious with *waltz*. So one way of trying to get around the symbol grounding problem is to assume that meanings relate to representations of our physical interactions with the world.

---

The view that cognitive activities such as understanding the meaning of something are bound up with representations of actual interactions with the world is part of the issue of **embodied cognition**. It is a short step from words to sentences. Sentences, on this view, suffer from the symbol grounding problem if they are not connected to perception and action. For instance, in order to understand the unconventional use of the word *elbow* in verb form (e.g. 'John *elbowed* the pencil to Mary'), we have to consider the range of actions of which the elbow is capable.

Is there any experimental evidence for bodily involvement in the understanding of entire sentences depicting simple actions? Recent work by Glenberg and Kaschak (2002) and their colleagues suggests that the direction of real movement underpins the comprehension of transfer and movement sentences.

### 7.1

### Research study

#### Is language understanding rooted in bodily movement?

Glenberg and Kaschak (2002) compared people's responses to two types of sentences, denoting motion either 'towards' or 'away', using imperative, physical transfer and abstract transfer examples:

'Towards' sentences:

Open the drawer (*imperative*)

Courtney handed you the notebook (*physical transfer*)

Liz told you a story (*abstract transfer*)

'Away' sentences:

Close the drawer (*imperative*)

You handed Courtney the notebook (*physical transfer*)

You told Liz a story (*abstract transfer*)

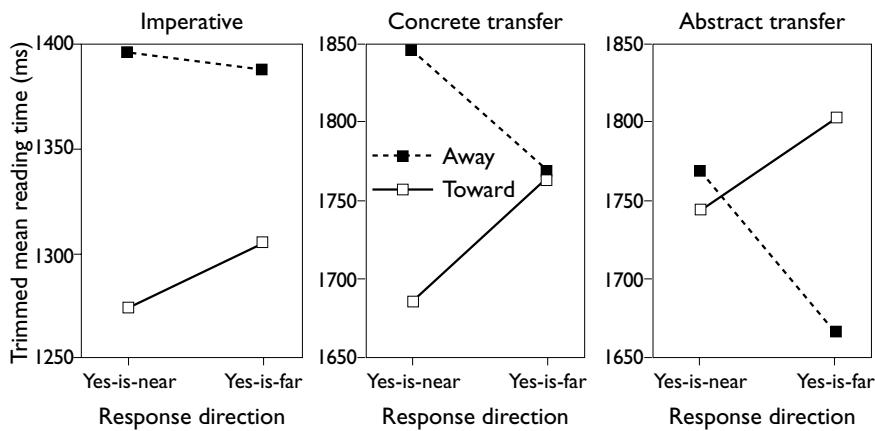
Participants were presented with sentences of these two types, along with nonsense sentences, like 'Boil the air', and asked to judge whether each sentence was 'sensible'. They were to indicate this by pressing a button that was either on the far end of a response box (i.e. away from their body), or at the near end. At the outset of each trial, the response finger rested on a centre button, so that to respond, participants actually had to make a movement either towards themselves or away.

The rationale was that when 'away' sentences are comprehended, part of the understanding involves a mental simulation of transferring the object (concrete or abstract) away from the body. Responses that involve a physical movement away from the body would be consistent with this mental simulation for the 'away' sentence, whereas a movement towards the body would conflict with that simulation. The prediction was that responses would be quicker when the understanding of the sentence was consistent with the movement required to make the response, and slower when these conflicted. That is, for 'away' sentences, a response that involves making an away movement (to the far end of the response box) would be quicker than a response involving a towards movement (to the near end of the response box). The opposite should apply to 'towards' sentences.

The results in Figure 7.1 show that for 'away' sentences, the 'yes-is-far' response (i.e. when people make a 'yes' response involving a movement away from their bodies to the far end of the response box) is quicker than the 'yes-is-near' response. For 'towards' sentences, the opposite is true – the 'yes-is-far' response is slower than the 'yes-is-near' response. Although the effects appear weaker for



the imperative sentences, Glenberg and Kaschak (2002) take these results as supporting their view that the understanding of transfer sentences is rooted in the actions underlying the transfers themselves.



**Figure 7.1** Time to initiate a response in the Glenberg and Kaschak (2002) study. Although there is some variability depending on the type of sentence used, there is good evidence for a compatibility effect between direction indicated by verb and direction of response

Source: Glenberg and Kaschak, 2002

## 2.2 Special topics in understanding text

Before moving on to language production and dialogue, we shall focus on two relatively new, emerging, and important issues: **shallow processing** and **perspective in communication**.

### 2.2.1 Shallow processing and selective processing

Just how completely do we utilize the meanings of words in establishing sentence meaning? Just how much detail goes into the representations of discourse we end up with after reading? Before reading on, quickly complete the questionnaire below on underspecification and depth of processing.

#### ACTIVITY 7.3

Complete the following questionnaire as quickly as possible (All questions apply to what is true in the United Kingdom).

- 1 Can a 16-year-old girl marry without her parent's permission?
- 2 Can a man marry his natural sister?
- 3 Can a person marry their first cousin?
- 4 Can a man marry his widow's sister?
- 5 Can a woman marry her uncle?

## COMMENT

These questions tap into your knowledge about what the marriage laws are. But they do more than that. One of the questions usually gets the answer 'yes' when in fact it has to be no: Can a man marry his widow's sister? The answer has to be 'no', because a person who has a widow is dead.

Anomalies such as these show how word meanings are not necessarily fully analysed and/or integrated into the mental representation of a discourse (which of these alternatives holds is an empirical issue). There are many such cases (Sanford and Sturt, 2002), the best known being the so-called Moses illusion (Erickson and Matheson, 1981):

- Moses put two of each sort of animal on the Ark. True or False? **Give your answer before reading on.**

The answer, of course, is that it is false because Noah was the one with the Ark.

---

The failure to use the full meaning of a word is a demonstration of shallow processing: not shallow in the sense of being sensory rather than to do with meaning, but rather in the sense of dealing only superficially with meaning. This can easily be seen with contrasts between different versions of the Moses illusion. For example, if *Adam* is substituted for *Moses*, then everyone spots that he didn't put any animals on the Ark (Van Oostendorp and De Mul, 1990). It is argued that this is because *Moses* is more similar in 'meaning' to *Noah* than is *Adam* (according to participants' ratings). So, it is not that people don't process the meanings of words in anomalies: rather, they do not process them very deeply. Barton and Sanford (1993) provide further evidence using the anomaly 'After an aircrash, where should the survivors be buried?'

In Chapter 6, and in Section 2.1.2 we saw evidence suggesting that word meaning is retrieved immediately a word is read. So, are these findings inconsistent with the incremental interpretation hypothesis? No – because, if a word is a really poor fit in meaning, then it is noticed: it is only when it is a close (but wrong) fit that there are problems. So meanings may still be used immediately, but only part of the full meaning may be used. For this reason, it may be best to say that readers immediately *initiate* meaning retrieval, but that this may be incomplete.

It is perhaps even more interesting that the extent to which meaning is processed depends upon the syntactic construction of sentences. Baker and Wagner (1987) presented participants with sentences like 16 and 17, and asked them to say whether the statements were true or not. Try the first one for yourself:

- 16 The liver, which is an organ found only in humans, is often damaged by heavy drinking.
- 17 The liver, which is often damaged by heavy drinking, is an organ found only in humans.

Participants spotted that sentences like 16 were false about 69 per cent of the time: here, the statement 'an organ found only in humans', which is of course false, is in the subordinate position – that is, in a clause that is subordinate to the main

sentence. In 17, where the statement is in the main clause of the sentence, errors were spotted 80 per cent of the time. So, putting information in a subordinate clause makes it less detectable than if it is in the main clause. Practically speaking, if you don't want people to scrutinize what you are saying too closely, then put the bit that you want them to miss in a subordinate position!

Subordination is an example of how syntax influences the extent of processing. There is also evidence that the part of the sentence on which focus, or emphasis, is put determines the depth of semantic processing. A **cleft sentence** has a structure like 'It was John who opened the door', the phrase *It was* being one half of the cleft. Using *It was* indicates clearly that the sentence answers the question 'Who opened the door?' With standard Moses illusion sentences, Bredart and Modolo (1988) showed that in 19 detection of the anomaly was much better than with 18:

- 18 Moses put two of each kind of animal on the Ark. True or False?
- 19 It was Moses who put two of each kind of animal on the Ark. True or False?

So, the focus of a sentence appears to receive deeper processing than the nonfocused elements. These simple observations open the way to developing processing theories in a new direction, by showing how the forms of sentences influence the amount of processing effort afforded the retrieval of meaning from words. There are many other situations in which shallow processing may occur, and establishing these will enable us to build more sophisticated accounts of language comprehension. Papers by Ferreira *et al.* (2002), and Sanford and Sturt (2002) illustrate the scale of this effect.

### 2.2.2 Perspective in communicating quantities

Language provides a point of view or perspective, and so controls the way we reason about things. When we read a novel, for instance, we typically take the perspective of the principal character. Perspective effects are found everywhere in language, and represent an important phenomenon for theories of understanding. However, some of the very simplest cases have practical consequences for us all, as we describe here.

For instance, there is a growing interest in how to communicate everyday risks more effectively. For example, medicines may have side effects, and how these are described influences our perception of the risks involved. Similarly, descriptions of foodstuffs may be slanted to make the foodstuff sound as healthy as possible. Recent work from several investigators poses some interesting challenges, both practical and theoretical.

Look at the different ways the fat content of food might be portrayed to a consumer:

- contains 9% fat
- contains less than 10% fat
- is 90% fat free

If you were trying to sell a product, which description would be best? By far the most common formulation is *% fat free*. Levin and Gaeth (1988) found that describing minced beef as *75% lean* rather than *25% fat* led people to rate the beef as leaner, less

greasy, and of higher quality, an effect that lasted even after they had tasted the beef! In much the same way, the fat-free formulation draws attention to leanness, while the *%fat* formulation indicates that there is fat content. Sanford *et al.* (2002) studied the basis of this phenomenon. They noted that, assuming people think that fat is unhealthy, the formulations lead to different evaluations. So, intuitively, in 20, the phrase *which is a bad thing* provides an intuitively acceptable completion of the sentence, whereas in 21, *which is a good thing* provides the acceptable completion (even though both initial clauses depict the same amount of fat):

- 20 This product contains 10% fat, which is a bad thing.  
 21 This product is 90% fat free, which is a good thing.

In a reading time experiment, participants saw materials like:

A new home-made style yoghurt is to be sold in supermarkets.  
 The yoghurt [contains 5% fat/25% fat]/ [is 95% fat free/is 75% fat free].  
 It is widely believed to be a healthy/unhealthy product.

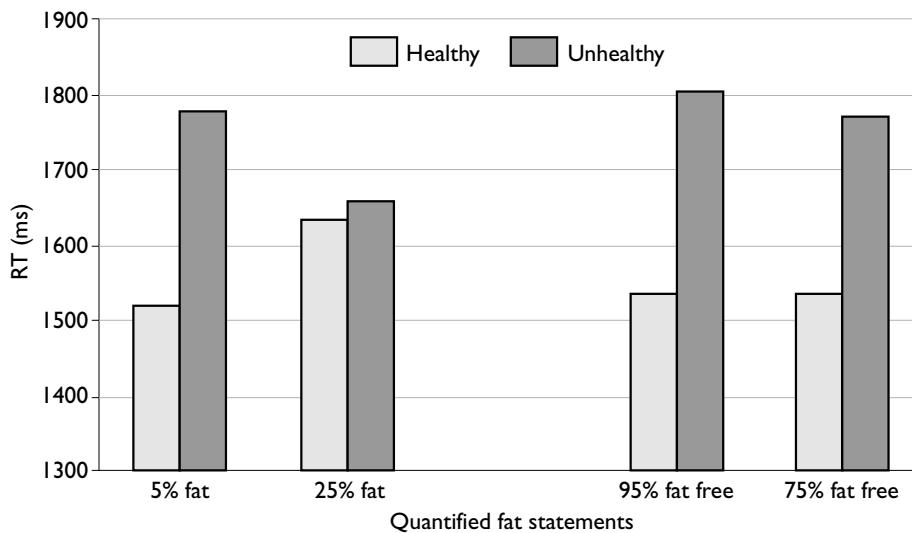
The brackets and slashes indicate alternative options for different conditions of the experiment. So, for instance, the *5% fat* formulation in the fat statement sentence could be followed by either the *healthy*, or the *unhealthy* continuation in the next sentence.

Participants read texts like these, one sentence at a time, using a self-paced procedure. The prediction was that if a description makes a product sound healthy, then the ‘healthy product’ version of the final sentence should be read more quickly, than the ‘unhealthy product’ version, because it would make more sense and be easier to integrate. The results are shown in Figure 7.2. When the *%fat* formulation is used, *5% fat* leads to faster reading times for the healthy version. This pattern changes for *25% fat*: reading time for ‘healthy’ goes up, while reading time for ‘unhealthy’ goes down. So, *5% fat* is taken as healthier than *25% fat*, as one might expect. People are bringing their knowledge of expected amounts of fat to bear on the situation. But look at the results for fat free. Both *95% fat free* and *75% fat free* lead to faster integration of the ‘healthy product’ target, and, there is no difference between these two. So, given the fat-free formulation, people appear *not* to be using their knowledge. Sanford *et al.* (2002) suggested that the *fat-free* formulation effectively stops people from utilizing the kind of knowledge they would use with the *%fat* formulation.

These are very clear effects of perspective: in the case of *% fat free*, the perspective is on the amount of non-fat (or healthy) ingredient, whereas in the case of *% fat*, the perspective is on the amount of fat. Although these two formulations actually depict the same amount of fat, they lead to quite different mental operations during understanding.

Perspective effects like this are subtle, but can influence how we think of things. With risks, for instance, compare the following formulations:

- Side effects, including headaches, occur rarely.
- Side effects, including headaches, occur occasionally.



**Figure 7.2** Mean reading times (RTs) for the final sentences. ‘Healthy’ and ‘unhealthy’ refer to choice of word in final sentences

Source: Sanford et al., 2002

Although *rarely* and *occasionally* both denote headaches occurring a small, unspecified proportion of the time, you could fit the continuation *which is a good thing* to the first one, and *which is a bad thing* to the second one. They point to different perspectives, so that the same chance of a side effect can sound *good*, or *bad*.

## Summary of Section 2

---

In Section 2, we have seen how the way language is used can affect our understanding of what is written. Indeed, given language is such an important tool for communication, it is unsurprising that language use has such an impact on such things as our judgements of risk and our ability to understand and solve basic problems.

- Establishing the coherence of a text involves more than merely combining the literal meanings of the words it contains. Examples of this include resolving anaphoric reference, deriving non-literal meaning, drawing inferences, drawing on world knowledge, and the role of the embodiment of actions and perceptions in understanding.
- The meanings of words are not always fully processed, and the depth of processing depends on focus.
- The understanding of written language we derive depends in part on the perspective or point of view provided by the text.

### 3 Language production as a self-contained process

So far we have considered only language comprehension, thus reflecting the history of psycholinguistics, which for many years treated language processing as equivalent to language comprehension. However, the primary setting for language use is in dialogue, and dialogue highlights the importance of language production. Since the 1990s, there has been a growing interest in production and even more recently in the dynamics of dialogue. So the remainder of this chapter concentrates on language in action, first in relation to production as a self-contained process, and then in relation to both production and comprehension as they occur in dialogue.

For any competent speaker, language production seems a straightforward process. For instance, when holding a conversation you are rarely aware of encountering any difficulty in formulating your utterances. However, the apparent ease of language production in informal settings, such as during a conversation, disguises the fact that it is a complex multi-stage process. The complexity is more apparent when producing a monologue (e.g. giving a talk or a presentation). For instance, imagine that you have to give a talk about this chapter of the book. Suddenly, language production becomes difficult. You may have trouble finding the right words to express yourself, or organizing what you want to say in a readily understandable form; you might even have trouble producing strictly grammatical sentences.

Here, we introduce the language production process in both these stages. First, we consider production as a self-contained process, as when you have to produce something like a talk. We look at speech errors and what they can tell us about the organization of production system. We then look at two special aspects of production: how speakers design their utterances and how speakers monitor their own speech. Second, we turn to dialogue and consider why language production is more straightforward in the informal setting of dialogue than it is with monologue.

#### 3.1 Speech errors and the architecture of the language production system

Much of what is known about language production has come from the study of speech errors. So first we consider what speech errors can tell us about the overall organization of the language production process and then look in more detail at recent work on two particular topics – first, how speakers design their utterances for particular listeners and second, how speakers monitor their spoken output.

Speakers make relatively few errors in normal speech (roughly 1 in every 2,000 utterances contains an error), but the errors they do make provide useful evidence about the overall organization of language production system. Table 7.2 shows the range of different kinds of speech errors that have been regularly observed.

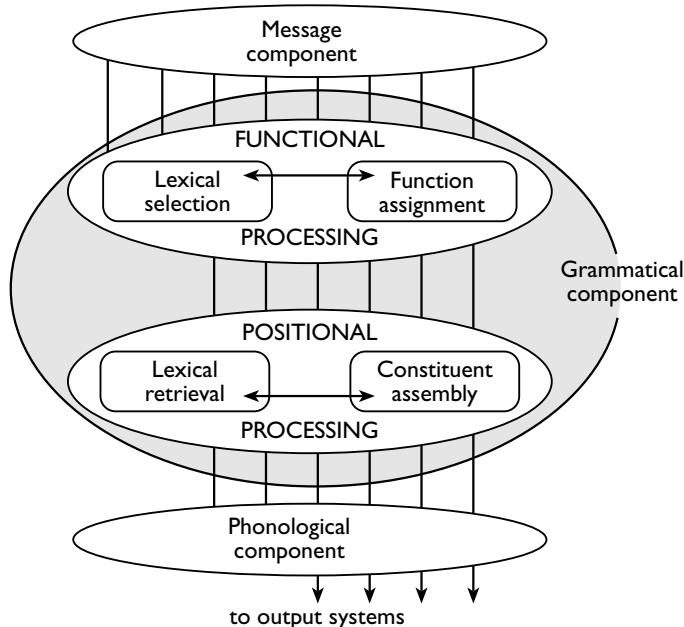
**Table 7.2** Sample speech errors

Type of error	Intended utterance	Error
1 Word anticipation	bury me right with him	bury him right with him
2 Sound anticipation	the lush list	the lust list
3 Word perseveration	evidence brought to bear on representational theories	evidence brought to bear on representational evidence
4 Sound perseveration	President Bush's budget	President Bush's boodget
5 Word exchange	the head of a pin	the pin of a head
6 Sound exchange	occipital activity	accipital octivity
7 Stranding exchange	the dome doesn't have any windows	the window doesn't have any domes
8 Phrase exchange	the death of his son from leukaemia	the death of leukaemia from his son
9 Semantically related word-substitution	I like berries with my cereal	I like berries with my fruit
10 Phonologically related word-substitution	part of a community	part of a committee
11 Sound substitution	the disparity	the disparig
12 Word blend	it really stood/stuck out	it really stook out
13 Phrase blend	at large/on the loose	at the loose

Source: Bock and Huitema, 1999

At first sight it may seem as if almost any kind of error can occur, but closer examination reveals interesting limitations. Take, for example, the exchange errors (5, 6, 7 and 8). It turns out that exchange errors always occur between items of the same syntactic category: nouns exchange with other nouns, verbs with other verbs; and when phonemes (minimal units of speech sound) are exchanged it tends to be consonants with consonants and vowels with vowels. This immediately suggests that the choice of the linguistic units in formulating the utterance (e.g. choice of words or choice of phonemes) is distinct from the grammatical formulation of the utterance in terms of the ordering of those units. In other words, the choice of the word or phoneme occurs at a separate stage from the decision about where the word or phoneme should be placed in the utterance sequence. Otherwise, it would be difficult to explain how the units could end up so far away from where they were supposed to be. Another striking phenomenon is what is called ‘stranding’. Take Example 7 in Table 7.2. The speaker had intended to say ‘The dome doesn't have any windows’, but the plural *windows* was placed in the part of the sentence where the singular word *dome* should have been and vice versa. What came out was not ‘The *windows* doesn't have any *dome*’ but rather ‘The *window* doesn't have any *domes*’. Hence, the assignment of the number feature on the words *window* and *dome* (i.e. whether they were marked as singular or plural) seems to involve a separate stage in the process from the choice of word and placement of that word in the sentence. Another such example of stranding is when the speaker intended to say ‘If that was done to me’ and it came out as ‘If *I* was done to *that*’ and not

'If *me* was done to that'. In this case, it is the grammatical case marking of the pronoun (i.e. *I* for subject of the verb was *done* and *me* for object) that is assigned separately after the pronoun has been put into that position in the sentence. So, the examples indicate that speech errors are more subtle than at first they seemed. Also, they tell us something about what sorts of operations go together in producing an utterance.



**Figure 7.3** A summary of the organization of the language production system

Source: Bock, 1995

On the basis of such patterns of errors, the overall organization of the language production system is commonly viewed as in Figure 7.3 (Bock, 1996 and Levelt, 1989 offer fuller accounts). It has three main processes. First, there is the process of formulating the message in a prelinguistic form (the message component). Then, there are two distinct processing systems: the first concerned with formulating the grammatical aspects of the utterance (the grammatical component) and the second (the phonological component) is concerned with converting this into the appropriate sequence of sounds. There are many reasons for making this split. First, speech errors are predominately either lexical errors or errors involving phonemes. In fact, the first six kinds of error in Table 7.2 can either be lexical (e.g. 1, 3 and 5) or phonological (e.g. 2, 4 and 6), but rarely involve other units of speech. In other words, even though morphemes (i.e. meaningful bits or words like the *meaning* and the *-ful* in *meaningful*) are more prevalent than words, there are very few morphemic errors. Similarly, errors involving the more common phonetic features (i.e. sound segments out of which the phonemes are made) are much rarer than errors involving phonemes. This suggests that the two main processes either arrange words (grammatical encoding) or arrange phonemes (phonological encoding).

Another piece of evidence relates to the earlier point about lexical exchanges in which nouns exchange with nouns and verbs with verbs. If grammatical and phonological encoding occurred together we would expect sound exchanges to occur within words of the same grammatical class. However, it turns out that the grammatical category of the words in which sound exchanges occur can be completely variable (e.g. *occipital* is an adjective and *activity* is a noun). In other words, grammatical class is only relevant to lexical exchange errors and not to phonological exchange errors. Finally, it turns out that the different kinds of exchange error, lexical and phonological, occur across spans of quite different length. Phonological exchanges tend to be from adjacent words whereas lexical exchanges tend to occur across adjacent phrases. This would suggest that the two kinds of process – grammatical and phonological – operate across rather different domains. Whereas grammatical processing can take a long-term view of the utterance, phonological processing only operates locally one word or two words at a time.

### ACTIVITY 7.4

---

Now try repeating the words below quickly (this tongue-twister exercise is based on Wilshire, 1999). Start by repeating pod-cab- etc., then after a while have a go with moss-knife- etc. If you can, record your repetitions on tape and then examine them for different kinds of phonological speech error.

**POD CAB CORD PUB**

**MOSS KNIFE NOOSE MUFF**

### COMMENT

You probably noticed that you started to make errors that involved substituting the speech sounds from one word into the next. Typically, the errors were anticipations where you might have said *nuss* for *noose* anticipating the following *muff*. This exercise illustrates how phonemes can be substituted in adjacent words during normal speech.

---

So there is a basic distinction between the grammatical and phonological encoding processes. Looking in more detail at the model we can see that each of these also involves distinct operations. For example, in grammatical processing there are operations that select the word to be uttered (lexical selection) and processes that determine its semantic function in the sentence (function assignment). So, for example, if it is a noun, whether it is to be the agent of the verb, the person that carries out the action, or the patient, the person acted upon. Still within the grammatical component, there is another process that recovers the word form (lexical retrieval) and another that builds up the grammatical constituents of the utterance (constituent assembly). Again, evidence for this distinction can be found in speech errors. Stranding errors, such as Example 7 in Table 7.2, indicate that word selection occurs separately from retrieving the precise form of the word. In ‘The window doesn’t have any domes’, an abstract representation of the word *window* is selected and put into the sequence before its precise form as *window* is retrieved, hence it can end up as *window* rather than the intended *windows*.

At the next level down in the system, the phonological component, it is sometimes claimed that there are different sub-components to deal with phonemes, on the one hand, and larger syllabic units that carry the stress patterns in the speech on the other (Levelt, 1989). We could go into much more detail about the organization of the language production system, but it is beyond the scope of a general cognition course. Instead, let's consider a couple of topics in more detail: message selection and audience design.

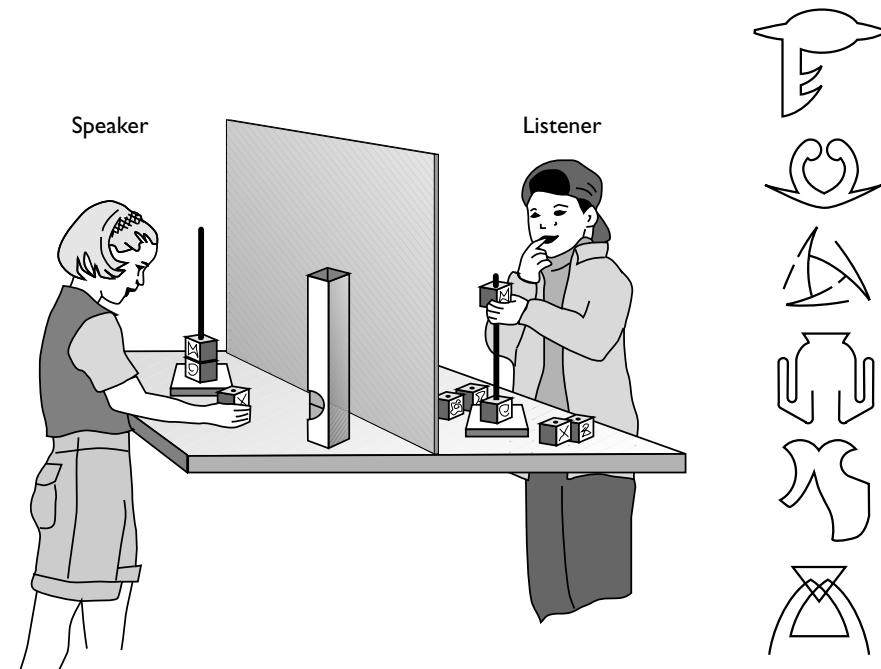
### 3.2 Message selection and audience design

In looking at the language production process we have adopted a similar model to that assumed in the work on language comprehension you met earlier in Chapter 6. Basically, production when viewed as an isolated process is seen as a kind of construction process from an idea via intermediate levels to a sequence of articulated sounds. But of course when we use language to communicate, the speaker has to make a number of more complicated general decisions about how to formulate what he or she wants to say in such a way that it will make sense for that particular listener. This general topic is what has been called **audience design**.

Audience design is an interesting part of the production process because it requires the speaker to draw complicated inferences about what the listener knows. These inferences are more complicated than you might imagine because they usually involve establishing what is called **common ground**. Technically, common ground relates to the knowledge that the speaker and listener share *and that they both know that they share* (Clark, 1996). Common ground is important because it affects how you should formulate your utterance in such a way that you can be sure it will be understood as you intended it to be. For example, say you are going around an art gallery with a companion and you turn to gaze at a painting that you really like. You might say to them 'It's great isn't it'. Now under certain circumstances that would be a perfectly felicitous statement and convey to your companion that you really liked that particular painting. However, under other circumstances it would be totally uninformative for them. It all depends on what you know that they know at that time and what you know that they know that you know, and so on. For instance, if you can see from the corner of your eye that they are looking at the same painting and that they can see that you are also looking at that painting then the statement is quite clear. Both of you take the thing that is really great to be the painting that you both know that you are both looking at. However, on other occasions the statement may not be felicitous. Anything that blocks establishing common ground could lead to misunderstanding. For instance, if you were standing looking at the painting but there was a barrier obscuring your view of your companion, then they would have no basis for establishing what it was you intended to speak about in saying 'It's great, isn't it'. Even if the barrier were a one-way mirror allowing you to see them, but not allowing them to see you and to see what you were looking at, the statement would be infelicitous.

**7.2****Methods****The referential communication task**

In this task (see Figure 7.4) there is a director (speaker) and a matcher (listener) who are separated by a thin partition. The director picks wooden blocks impaled on a stake and has to communicate to the matcher the nonsense pattern shown on the block (see the right side of Figure 7.4 for examples). The matcher then chooses the appropriate block from his or her pile and then puts it onto their stake. With this task it is possible both to analyse what the communicators say to each other and to establish how accurately they can communicate the patterns by comparing the order of items that the director started out with to the order of items that the matcher ended up with. You could try it out with your friends.



**Figure 7.4** The referential communication task

Source: Glucksberg and Danks, 1975

This might seem to be a rather special and complicated example, but audience design enters into almost everything we say. Take for example an experiment by Isaacs and Clark (1987). They used the referential communication task shown in Figure 7.4 (Krauss and Weinheimer, 1967). However, in their case the director had to indicate to the matcher which picture he or she was looking at from a set of pictures of buildings in New York City. In effect, the director had to describe the picture unambiguously to the matcher. Now the twist in the experiment was that the communicating pairs were chosen so that either one or both or neither were New Yorkers. In other words, the degree of common ground between the different pairs could be quite different and the question was whether or not the directors would alter the design of their descriptions according to their assessment of common ground. It

turned out that everyone was extremely good at the task and almost immediately established whether they were both New Yorkers, only one was a New Yorker or neither was. In fact 85 per cent of pairs knew into which category they both fell after describing only two of the cards. But perhaps more interesting was how the speakers then adapted their descriptions according to whom it was that they were speaking. New Yorkers speaking to New Yorkers would typically just name the building in the picture (e.g. ‘Chrysler building’) whereas New Yorkers speaking to non-New Yorkers would describe the picture itself (e.g. they might say ‘Well it has three buildings in it, a tall one and two short ones’) and so design their utterances according to what they knew was the common ground.

This experiment illustrates nicely how language production in a real context involves much more than just translating ideas into sounds. It also requires a complex assessment of what the listener knows at the time – including what is in common ground between the speaker and listener. However, there is controversy over the extent to which speakers always take common ground into account. For example, Horton and Keysar (1996) found that speakers under time pressure did not produce descriptions that took advantage of what they knew about the listener’s view of the relevant scene. In other words, the descriptions were formulated with respect to the speaker’s current knowledge of the scene rather than with respect to the speaker and listener’s common ground. Similarly, Keysar *et al.* (1998) found that in visually searching for a referent for a description listeners are just as likely to initially look at things that are not part of the common ground as things that are. In other words, listeners also do not seem to always take advantage of common ground.

Nevertheless, Horton and Keysar (1996) found that with less time pressure, speakers often did take account of common ground in formulating their utterances, and Keysar *et al.* (1998) argued that listeners at a later monitoring stage take account of common ground in comprehension.

So, all in all, it seems that audience design and the extent to which the audience as comprehender is sensitive to design is a complicated issue. In the absence of time pressure, both language producers and language comprehenders are able to take into account their common ground in processing an utterance. However, when under time pressure, this kind of complex assessment of listener by speaker and vice versa is one of the first parts of the process to suffer. Below we consider how this separation of processes in production may depend upon the distinction between initial formulation of an utterance and subsequent monitoring and correction of that utterance.

### 3.3 Self-monitoring

An important part of the process of speech production is being able to monitor and correct what you are saying (Hartsuiker and Westenberg, 2000). We know that speakers are always doing this because natural speech is full of minor hesitations and dysfluencies in which the speaker briefly stops and corrects or repairs their utterance.

Now, there is a real issue as to how this monitoring process operates. In its most straightforward form, monitoring can work by the speaker listening to and comprehending their own output. Then, as soon as they encounter something that doesn’t match what they had originally intended to say they can stop the speech, reformulate the utterance and continue with the repaired fragment. For example, take

Utterance 6 in Table 7.3 (see Section 4.1 below). Here, the speaker starts out saying ‘The left’ but then realizes that he should have been more explicit. So he stops and restarts his utterance with the repair ‘going from left to right in the second box’. It is straightforward to assume the existence of such an ‘outer-loop’ monitoring process (monitoring based on speech output) because we know that we can perfectly well understand what we are saying to someone else. More controversial is the idea that monitoring can also operate at earlier stages in the production process. For instance, it can operate at the message formulation stage or later at the stage of phonological encoding. This is called ‘inner-loop’ monitoring, which is monitoring based on something available before the speaker has had a chance to listen to what they are saying.

Evidence for inner-loop monitoring comes from a variety of sources. The most colourful evidence is from a speech error elicitation experiment by Motley *et al.* (1982). They used a device for producing speech errors which was a bit like the tongue-twister elicitation procedure described in Activity 7.4. Participants had to repeat context sequences which were likely to produce errors in subsequent critical items, such as the pair ‘barn door’ which might be mispronounced as ‘darn bore’. However, they also included critical pairs of items that if mispronounced would lead to a taboo word (e.g. ‘tool kits’ as ‘cool \*\*\*’). The crucial question was whether speakers were as likely to come up with the taboo forms as they were to come up with non-taboo forms. Motley *et al.* found that the taboo errors were much less likely to occur than non-taboo errors. So they argued that the taboo errors must have been filtered out before the utterance had been articulated. Now it is difficult to understand how such a pre-articulatory filtering can occur without some form of ‘inner-loop’ monitoring.

Another kind of evidence comes from examining the temporal characteristics of speech errors and their corrections. The crucial measure is the time between producing the incorrect word and producing the repair (e.g. the time between the ‘the left ...’ and the ‘going from left ...’ in Utterance 6 in Table 7.3 below). Hartsuiker and Westenberg argued that any error correction time interval of less than 150 ms couldn’t reflect the outer-loop monitoring process, because this would not allow sufficient time to comprehend the output, reformulate it and then restart the utterance. Looking at several collections of such errors they found a high preponderance of these short latency restarts. In fact their results suggested that the majority of speech repairs were based on ‘inner-loop’ as opposed to ‘outer-loop’ monitoring. An additional interesting feature of these two kinds of monitoring is that whereas outer-loop monitoring appears to tax attention, ‘inner-loop’ monitoring does not. This explains why when people speak under time pressure they tend to produce less overt speech repairs than when speaking more slowly and the overt repairs tend to involve short latency restarts. It is assumed that, with high time pressure, monitoring and repair shift even more in favour of the low attention ‘inner-loop’ route. It is also one reason why Keysar *et al.* (1998) argued that audience design might only influence production at a later stage on the basis of self-monitoring. The idea was that such monitoring would be associated with the outer loop and so would be less effective when speaking under time pressure.

A similar distinction between ‘inner-’ and ‘outer-loop’ monitoring is made for monitoring of physical movements, such as when grasping something or when

picking something up (Blakemore *et al.*, 2002). Again, it seems that the inner-loop equivalent for motor control is not accessible to consciousness and does not tax attention to the same degree that the outer-loop system does.

## Summary of Section 3

---

Section 3.1 outlined some of the basic processes and architecture of language production. Sections 3.2 and 3.3 considered two special topics in language production that lie beyond this basic framework:

- The first topic, audience design, concerned ways in which the speaker attempts to take the listener into account during production. There is some controversy over the degree to which audience design, at least in relation to taking into account the common ground, is an obligatory part of the normal language production process. Clark and colleagues have argued that speakers aid their listeners by taking account of common ground when formulating their utterances. However, others have argued that common ground is only taken into account on the basis of an optional outer-loop monitoring and repair process.
- The second topic concerned the mechanisms of self-monitoring in speech. The important issue here is whether such monitoring can only occur on the basis of a speaker listening to himself or herself—outer-loop monitoring—or whether it also proceeds on the basis of prearticulatory monitoring—inner-loop monitoring. Evidence from the time course of self-repair strongly indicates the prevalence of prearticulatory inner-loop monitoring.

## 4 The challenge of dialogue

So far we have considered language production as if it were a process completely isolated from comprehension. However, language processing most often occurs in the context of a dialogue where each participant both produces and comprehends more or less at the same time. How does this kind of interaction affect the production and comprehension process?

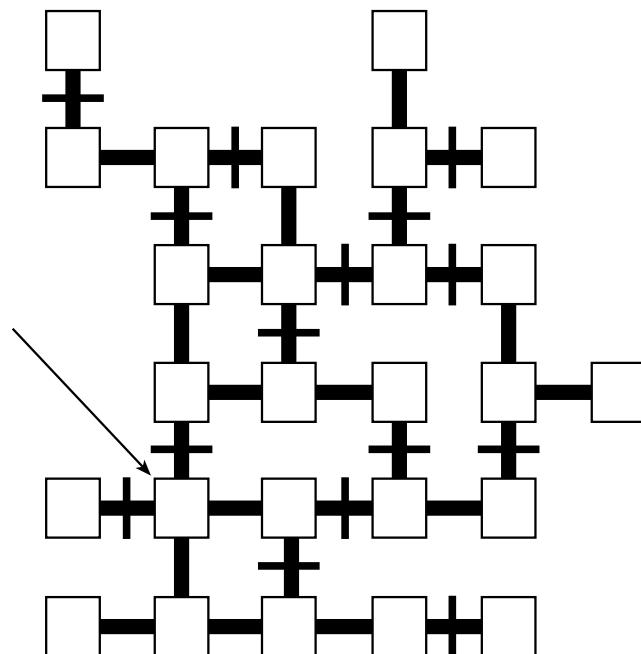
First, we take a look at what happens in dialogue and how the language used is different in dialogue from monologue. This then leads to a more general discussion about how dialogue and monologue involve different kinds of processing and in turn how this influences the nature of production and comprehension in a dialogue context. Finally, we consider a recent model of language processing in dialogue that takes these differences into account.

### 4.1 What is dialogue?

The example in Table 7.3 overleaf comes from a transcript of two players in a cooperative maze game where one player *A* is trying to describe his position to his partner *B* who is viewing the same maze on a computer screen in another room.

**Table 7.3** An excerpt of dialogue, from Garrod and Anderson, 1987 (the position being described in all the utterances shown in bold is illustrated schematically in Figure 7.5)

- |    |   |
|----|---|
| 1  | B: Tell me where you are?   |
| 2  | A: Ehm: Oh God (laughs)   |
| 3  | B: (laughs)   |
| 4  | <b>A: Right: two along from the bottom one up:</b>                    |
| 5  | B: Two along from the bottom, which side?                             |
| 6  | A: The left: going from left to right in the second box.              |
| 7  | B: You're in the second box.  |
| 8  | A: One up: (1 sec.) I take it we've got identical mazes?              |
| 9  | B: Yeah well: right, starting from the left, <b>you're one along:</b> |
| 10 | A: Uh-huh:  |
| 11 | B: <b>and one up?</b>   |
| 12 | A: Yeah, and I'm trying to get to ... etc.                            |



**Figure 7.5** The arrow points to the position on the maze that A and B are trying to describe in the dialogue extract shown in Table 7.3. Notice that the two descriptions in the text are in fact different – *Two along from the bottom one up* vs. *One along ... one up*

At first glance the language looks disorganized. Strictly speaking many of the utterances are not grammatical sentences – only one of the first six contains a verb. There are occasions when production of the same sentence is shared between the speakers, as in Utterances 7–8.

In fact the sequence is quite orderly so long as we assume that dialogue is a joint activity (Clark, 1996). In other words, dialogue involves cooperation between

interlocutors in a way that allows them to sufficiently understand the meaning of the dialogue as a whole; and this meaning results from these joint processes. So, dialogue is orderly to the extent that it requires coordination to establish consensus between the two speakers.

## 4.2 Dialogue and consensus

In a piece of written text, whether it is a newspaper article or the chapter of a learned volume, the meaning is there on the page waiting to be extracted. If it is well written and you are a competent reader, then you should be able to come to an interpretation which matches roughly what the writer intended to convey. However, this does not depend on establishing any kind of consensus with the author. After all, he or she may well be long dead and gone.

In dialogue the situation is very different. Dialogue is organized around establishing consensus. First, dialogue turns are linked across interlocutors (Schegloff and Sacks, 1973). In Table 7.3 opposite, an imperative question, such as 1, ‘Tell me where you are?’ calls for a response, such as 4, ‘Right: two along from the bottom one up’. Even a statement like 4 cannot stand alone. It requires either an affirmation or some form of query, such as 5, ‘Two along from the bottom, which side?’ This means that production and comprehension processes become coupled. *B* produces an imperative question and expects an answer of a particular type; *A* hears the question and has to produce an answer of that type. For example, after saying ‘Tell me where you are?’ in 1, *B* has to understand ‘two along from the bottom one up’ in 4 as a reference to *A*’s position on the maze; any other interpretation is ruled out.

Second, the meaning of what is being communicated depends on the interlocutors’ agreement or consensus rather than on dictionary meanings and is therefore subject to negotiation. Take for example Utterances 4–11 in Table 7.3. In Utterance 4, *A* describes his position as ‘Two along from the bottom one up’, but the final interpretation is only established at the end of the first exchange when consensus is reached on a rather different description by *B* (9–11) ‘You’re one along ... and one up?’

So, in dialogue, the interpretation depends upon taking part in the interaction itself. This was nicely demonstrated in an experiment by Schober and Clark (1989). They used an experimental set-up similar to that used by Krauss and Weinheimer (see Figure 7.4) in which a director had to describe a sequence of abstract Chinese Tangram patterns to a matcher on the other side of a screen. However, in this experiment there was also a third person who overheard everything that was said but could not interact with the director. The overhearer had to try to pick the cards being described by the director in the same way that the matcher had to pick them. It turned out that overhearers, who could not interact with the director, performed consistently less well than the matchers who *could* interact with the directors. Schober and Clark argued that overhearers are at a disadvantage because they cannot control what the director is saying in the way that the participant can (e.g. a participant can always query what they fail to understand whereas an overhearer cannot). So hearing everything that is being said does not lead to a full understanding when you cannot interact directly with the speaker.

The third way that dialogue involves coordinated processing relates to the general problem of ambiguity discussed in Chapter 6. In the extract in Table 7.3, the participants spend a lot of time trying to work out a mutually acceptable and unambiguous description for *A*'s location on the maze. As we shall see below, this is achieved through a process of coordinating outputs with inputs: speakers always attempt to generate utterances that correspond semantically to the utterances which they have recently had to comprehend. For example, consider Utterances 4 and 5 in which *B* echoes the description 'Two along from the bottom'. As a result of such output–input coordination, the same expression comes to take on the same precise meaning within any stretch of dialogue.

Finally, dialogue participants try to establish a coordinated conception of their topic. In the case of the maze game illustrated in Figure 7.5 this amounts to converging on a common spatial concept of the maze's configuration. Thus, some people playing this game will refer to their locations by reference to *right-turn indicators*, *upside-down T-shapes* or *Ls on their sides*. These speakers, unlike the pair responsible for the dialogue illustrated above, conceive of the maze as a conglomeration of patterns or shapes each with a different name. Conversational partners often establish quite idiosyncratic conceptions of the topic (as with the use of *right-turn indicator* or *upside-down T-shapes*) but in well-managed dialogues they always align on the same idiosyncratic conception. Again, this process supports consensus, which is the fundamental goal of dialogue.

One of the reasons why dialogue presents such a challenge to processing accounts is that these interactive characteristics are difficult to reconcile with the standard view of communication as a one-way process of *information transfer*. And it is just such a view that underpins much of the work in psycholinguistics (and similarly much work described in Chapter 6). Here we argue that a more useful processing framework for dialogue may be based on the notion of *interactive alignment*. According to this account, dialogue participants come to align their linguistic representations at many levels. The alignment process helps them to come to a mutually satisfactory interpretation of what is being said and it greatly simplifies the basic processes of production and comprehension during dialogue.

### 4.3 A model of dialogue processing

The interactive alignment account starts with the simple observation that dialogue participants alternate between *speaking* and *comprehending*. Furthermore, the representations that are used for comprehension (whether they are syntactic, lexical or at the level of articulation) will activate or prime matching representations in production.

If we assume that representations active during comprehension remain active during subsequent production, then there will always be a tendency for interlocutors to coordinate outputs (productions) with inputs (what has just been understood). If we put two such systems together in a dialogue, then the overall system will only be completely stable if the two adopt aligned linguistic representations at every level. Pickering and Garrod (in press) go on to argue that this kind of interactive alignment of representations supports mutual understanding because alignment does not just occur within independent levels of the system but it also serves to link those levels with each other. In other words, the automatic alignment of representations at all

levels will tend to establish a kind of common ground between the two communicators which aids mutual understanding.

First, we look at the evidence for representational alignment during dialogue and consider how it may support the interpretation process. Then we go into a little more detail on one consequence of representational alignment in relation to language processing – what has been called routinization.

#### 4.3.1 Evidence for representational alignment

Dialogue transcripts are full of repeated linguistic elements and structures indicating alignment at various levels (Aijmer, 1996). Alignment of lexical processing during dialogue was specifically demonstrated by Garrod and Anderson (1987) and by Clark and colleagues (Brennan and Clark, 1996; Wilkes-Gibbs and Clark, 1992). These latter studies show that interlocutors tend to develop the same set of expressions to refer to particular objects and that the expressions become shorter and more similar on repetition with the same interlocutor, but are modified if the interlocutor changes.

Levett and Kelter (1982) found that speakers tended to reply to ‘What time do you close?’ or ‘At what time do you close’ (in Dutch) with a congruent answer (e.g. ‘Five o’clock’ or ‘At five o’clock’). This alignment may be syntactic (repetition of phrasal categories) or lexical (repetition of *At*). Branigan *et al.* (2000) found clear evidence for syntactic alignment in dialogue. Participants took it in turns to describe pictures to each other (and to find the appropriate picture in an array). One speaker was actually a confederate of the experimenter and produced scripted responses, such as ‘the cowboy offering the banana to the robber’ or ‘the cowboy offering the robber the banana.’ The syntactic structure of the confederate’s description influenced the syntactic structure of the experimental subject’s description and it did so much more strongly than in a comparable non-dialogue situation.

Alignment also occurs at the level of articulation. It has long been known that as speakers repeat expressions, articulation becomes increasingly reduced (i.e. the expressions are shortened and become more difficult to recognize when heard in isolation; Fowler and Housum, 1987). However, Bard *et al.* (2000) found that reduction was just as extreme when the repetition was by a different speaker in the dialogue as it was when the repetition was by the original speaker. In other words, whatever is happening to the speaker’s articulatory representations is also happening to their interlocutor’s. So the two representations are becoming aligned. There is also evidence that interlocutors align accent and speech rate (Giles *et al.*, 1992; Giles and Powesland, 1975).

Taken together, these findings indicate that something rather special happens when we process language in a dialogue setting. The representations called upon in production are already in some sense available to the speaker from his or her comprehension of the prior dialogue. Apart from helping the interlocutors to come to a truly aligned interpretation of what the dialogue is about, it also simplifies the production and comprehension processes themselves. One of the ways in which this may happen is through what has been called **routinization**.

### 4.3.2 Routinization in dialogue processing

The process of alignment means that interlocutors draw upon representations that have been developed during the dialogue. Thus, it is not always necessary to construct representations that are used in production or comprehension from scratch. One particularly important implication is that interlocutors develop and use routines (set expressions) during a particular interaction.

A *routine* is an expression that is ‘fixed’ to a relatively great extent. First, the expression has a much higher frequency in the interaction than the frequency of its component words would lead us to expect (i.e. the combination of words occurs more often in the dialogue than in the language in general). Second, it has a particular analysis at each level of linguistic representation. Thus, it has a particular meaning, a particular syntactic analysis, a particular pragmatic use, and often particular phonological characteristics (e.g. a fixed intonation). Extreme examples of routines include repetitive conversational patterns such as ‘*How do you do?*’ and ‘*Thank you very much*’. Routines are highly frequent in dialogue. It has been estimated that up to 70 per cent of words in a standard dialogue occur as part of recurrent word combinations. However, different expressions can be routines to different degrees, so actual estimates of their frequency are somewhat arbitrary. Some routines are idioms, but not all (e.g. *I love you* is a routine with a literal interpretation in the best relationships).

Most discussion of routines focuses on phrases whose status as a routine is pretty stable. However, Pickering and Garrod (in press) also claim that routines are set up ‘on the fly’ during dialogue as a result of the interactive alignment process. They called this *routinization* and it represents one of the rather special features of language processing in a dialogue as opposed to a monologue setting.

## Summary of Section 4

---

In this section we have seen how language processing in dialogue may be rather different from language processing in monologue.

- Language processing in dialogue depends upon coordinated processes of production and comprehension, as in answering a question.
- Language processing in dialogue seems to involve direct participation from both interlocutors in creating a common understanding of the message. Hence, overhearers cannot fully understand what is being said in dialogue.
- Both production and comprehension in dialogue may be governed by an interactive alignment process that leads to routinization.

## 5 The monologue/dialogue distinction and group decision making

We opened this chapter by contrasting language use in the context of monologue and dialogue. But what are the consequences of this distinction beyond language processing itself? One interesting consequence relates to group decision making.

Communication is crucial to group decision making, whether it is a family deciding to move to a larger house, a parliament deciding on new legislation or a jury coming to a verdict. It is crucial because there is no other way in which a group can come to a consensus. A group decision, at least in its purest form, depends upon consensus. Remember, one of the main distinctions between dialogue and monologue is the way in which dialogue promotes consensus whereas monologue does not. So it is an interesting question as to what kind of communication processes operate within a group and how they might affect the way in which people are influenced by other members of the group in coming to a decision.

Imagine that you are a member of a committee discussing some particular issue at your work. Sometimes you will be aware of being highly engaged in discussion with just one or two other members of the committee; it is like a two-party dialogue. At other times you just sit back and listen to what the most vociferous member of that committee is saying. Now, afterwards, you happen to bump into someone who had been there and to your surprise you discover that what you thought was the crucial decision is not quite the same as what they thought was crucial. Typically, people's views about such things vary quite a lot. The question is, what affects those views and how does that relate to the communication process during the meeting itself. Are you going to agree more with the people that you had the interactive discussion with at the meeting or are you going to be influenced most by the vociferous and dominant member of the group?

The monologue and dialogue models of communication bear differently on this question. According to the information transfer or monologue account, a group discussion can be thought of as a process in which there are a series of monologues in which the current speaker broadcasts information to the rest of the group. Hence, you should tend to be influenced most by the person who says the most. Whom you speak next to in the discussion should have no special influence on your views. The interactive alignment or dialogue account makes a very different prediction. The people who should have most influence on your views are those with whom you directly interacted and there should be no particular reason why the dominant speaker should influence you most.

Fay *et al.* (2000) report a study that shows that both of these views are correct, but which applies depends on the size of the group holding the discussion. To test this they had two sizes of groups of students imagine that they were a university disciplinary committee who had to sit down and decide as a group what to do about a complex case of student plagiarism. First, each member of the group read a one-page description of the case and then, before discussing it, they each had to rank 14 relevant issues in terms of how important they felt they were to this case. The issues ranged from clearly relevant ones, such as the severity of the plagiarism, to more ambiguous issues such as the university's responsibility to the student. The groups then discussed the case for about 20 minutes and, after the discussion, each person again ranked the issues but now in terms of how important they thought they had been to the group as a whole. By comparing the agreement in the ranking scores between each member of the group after the meeting (after accounting for their pre-meeting agreements) it is possible to determine who has the strongest influence on whom with respect to the 14 key issues. Fay *et al.* then used the transcriptions of the discussions to establish:

- 1 Which members of the groups had either served as high-interaction partners for each group member and which had served as low-interaction partners.
- 2 Which members of each group had been dominant speakers as opposed to non-dominant speakers.

(For 1 they defined a high-interaction partner as a person who was most likely to speak either immediately before or immediately after each member of the group.)

They were then able to examine the degree to which everyone in the groups had been influenced either by their high-interaction partners or by the dominant speakers in the discussion.

The results were clear and quite striking. It turned out that in small group meetings with five members, people were only especially influenced by their high-interaction partners. There was no additional influence coming from dominant speakers. This is exactly what is predicted by the alignment account because (1) interactants automatically align with each other, and (2) an overhearer or side participant is not going to be influenced by others' interactions even when they involve the dominant speaker. However, in larger groups of 10, exactly the opposite pattern emerged. People in the larger groups were all influenced by the dominant speaker and there was absolutely no effect of interaction. This was as predicted by the monologue or information transfer account. Additional analysis of the meeting transcripts also supported the idea that in the small groups the utterances and turn pattern was just like the pattern in a two-party conversation. The utterances were shorter, there were many more interruptions and the pattern of speaker turns tended to conform to an *ABABA* pattern with the same two speakers taking alternate turns for extended periods of discussion.

## Summary of Section 5

---

Communication is critical to group decision making. Fay *et al.* found that decision making was influenced by communication within the group:

- Members of small groups were influenced most by their high-interaction partners.
- People in larger groups were all influenced by the dominant speaker.

## 6 Summary

We began this chapter by arguing that the simple interpretation account of language processing, whether in terms of comprehension or production, could not capture everything that happens when we use our language to communicate. Language in action involves access to general knowledge, inference beyond what is actually said and, in the case of dialogue, coordinated action. Without these additional processes, communication would fail. In fact, it is generally thought that much of the individual variation in reading ability, which is so relevant in school or university, stems from differences in readers' abilities to access the appropriate knowledge and hence their ability to integrate information in the texts they are reading (Garrod and Daneman, 2003).

In Section 2 we considered these additional processes in relation to text comprehension. We concentrated on the three basic issues that figure in most accounts of comprehension beyond the word: anaphora resolution, non-literal meaning and text inference. We then went on to discuss in more detail some of the hot topics in the area. So we looked at the argument for embodied representations of meaning, the problem of shallow or incomplete processing of text and the way in which perspective affects interpretation. In all these cases language comprehension involved more than just the translation of sounds or written symbols into meanings.

In Section 3 we turned our attention to language production. Language production when viewed as an isolated process seems to involve the same sort of interpretation processes assumed for comprehension in Chapter 6. We looked at how an examination of speech errors can help us to construct a provisional model of the language production system and at how issues such as audience design and self-monitoring show that speakers must consider non-linguistic discourse related factors when assembling utterances. However, when production and comprehension were considered in the context of dialogue, as we did in Section 4, these processes took on a different character. The important thing in dialogue processing is how production and comprehension processes become coupled to each other to produce aligned linguistic representations at every level. In turn, we saw how the contrast between language processing in monologue and dialogue had interesting consequences for such apparently non-linguistic activities as group interaction and decision making.

## Further reading

- Bock, K. (1995) ‘Sentence production: from mind to mouth’, in Miller, J.L. and Eimas, P.D. (eds) *Handbook of Perception and Cognition, Vol. 11, Speech, Language, and Communication*, Orlando, FL, Academic Press.
- Clark, H.H. (1996) *Using Language*, Cambridge, Cambridge University Press.
- Sanford, A.J. (1999) ‘Word meaning and discourse processing’, in Garrod, S. and Pickering, M.J. (eds) *Language Processing*, Hove, Psychology Press.

## References

- Aijmer, K. (1996) *Conversational Routines in English: Convention and Creativity*, London and New York, Longman.
- Baker, L. and Wagner, J.L. (1987) ‘Evaluating information for truthfulness: the effects of logical subordination’, *Memory and Cognition*, vol.15, no.3, pp.247–55.
- Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G. and Newlands, A. (2000) ‘Controlling the intelligibility of referring expressions in dialogue’, *Journal of Memory and Language*, vol.42, no.1, pp.1–22.
- Barton, S. and Sanford, A.J. (1993) ‘A case-study of pragmatic anomaly-detection: relevance-driven cohesion patterns’, *Memory and Cognition*, vol.21, no.4, pp.477–87.
- Blakemore, S.-J., Wolpert, D.M. and Frith, C.D. (2002) ‘Abnormalities in the awareness of action’, *Trends in Cognitive Science*, vol.6, no.6, pp.237–42.

- Bock, K. (1995) ‘Sentence production: from mind to mouth’, in Miller, J.L. and Eimas, P.D. (eds) *Handbook of Perception and Cognition: Vol.11, Speech, Language, and Communication*, Orlando, FL, Academic Press.
- Bock, K. (1996) ‘Language production: methods and methodologies’, *Psychonomic Bulletin and Review*, vol.3, no.4, pp.395–421.
- Bock, K. and Huitema, J. (1999) ‘Language production’, in Garrod, S. and Pickering, M.J. (eds) *Language Processing*, Hove, Psychology Press.
- Branigan, H.P., Pickering, M.J. and Cleland, A.A. (2000) ‘Syntactic coordination in dialogue’, *Cognition*, vol.75, no.2, B13–B25.
- Bredart, S. and Modolo, K. (1988) ‘Moses strikes again: focalization effects on a semantic illusion’, *Acta Psychologica*, vol.67, no.2, pp.135–44.
- Brennan, S.E. and Clark, H.H. (1996) ‘Conceptual pacts and lexical choice in conversation’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.22, no.6, pp.1482–93.
- Brown, G. and Yule, G. (1983) *Discourse Analysis*, Cambridge, Cambridge University Press.
- Clark, H.H. (1996) *Using Language*, Cambridge, Cambridge University Press.
- Erickson, T.A. and Matheson, M. (1981) ‘From words to meaning: a semantic illusion’, *Journal of Verbal Learning and Verbal Behavior*, vol.20, no.5, pp.540–62.
- Fay, N., Garrod, S. and Carletta, J. (2000) ‘Group discussion as interactive dialogue or as serial monologue: the influence of group size’, *Psychological Science*, vol.11, no.6, pp.481–6.
- Ferreira, F., Bailey, K.G.D. and Ferraro, V. (2002) ‘Good-enough representations in language comprehension’, *Current Directions in Psychological Science*, vol.11, no.1, pp.11–15.
- Fodor, J. (2000) ‘The mind doesn’t work that way’, Cambridge, MA, MIT Press.
- Fowler, C. and Housum, J. (1987) ‘Talker’s signalling of “new” and “old” words in speech and listener’s perception and use of the distinction’, *Journal of Memory and Language*, vol.26, no.5, pp.489–504.
- Garnham, A. (2001) *Mental Models and the Interpretation of Anaphora*, Hove, Psychology Press.
- Garrod, S. and Anderson, A. (1987) ‘Saying what you mean in dialogue: a study in conceptual and semantic co-ordination’, *Cognition*, vol.27, no.2, pp.181–218.
- Garrod, S. and Daneman, M. (2003) ‘Reading, the psychology of’, in Nadel, L. (ed.) *Encyclopedia of Cognitive Science*, 3, London and New York, Nature Publishing Group.
- Gibbs, R. (1983) ‘Do people always process the literal meanings of indirect requests?’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.9, no.3, pp.524–33.
- Giles, H., Coupland, N. and Coupland, J. (1992) ‘Accommodation theory: communication, context and consequences’, in Giles, H., Coupland, J. and Coupland, N. (eds) *Contexts of Accommodation*, Cambridge, Cambridge University Press.

- Giles, H. and Powesland, P.F. (1975) *Speech Styles and Social Evaluation*, New York, Academic Press.
- Glenberg, A.M. and Kaschak, M.P. (2002) ‘Grounding language in action’, *Psychonomic Bulletin and Review*, vol.9, no.3, pp.558–65.
- Glenberg, A.M. and Robertson, D.A. (2000) ‘Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning’, *Journal of Memory and Language*, vol.43, no.3, pp.379–401.
- Glucksberg, S. and Danks, J.H. (1975) *Experimental Psycholinguists: an introduction*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Glucksberg, S., Gildea, P. and Bookin, H. (1982) ‘On understanding nonliteral speech: can people ignore metaphors?’, *Journal of Verbal Learning and Verbal Behavior*, vol.21, no.1, pp.85–98.
- Glucksberg, S. and Keysar, B. (1990) ‘Understanding metaphorical comparisons: beyond similarity’, *Psychological Review*, vol.97, no.1, pp.3–18.
- Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*, London, Longman.
- Harnad, S. (1990) ‘The symbol grounding problem’, *Physica D*, no.42, pp.335–46.
- Hartsuiker, R.J. and Westenberg, C. (2000) ‘Word order priming in written and spoken sentence production’, *Cognition*, vol.75, no.2, B27–B39.
- Haviland, S.E. and Clark, H.H. (1974) ‘What’s new? Acquiring new information as a process in comprehension’, *Journal of Verbal Learning and Verbal Behavior*, vol.13, no.5, pp.512–21.
- Horton, W.S. and Keysar, B. (1996) ‘When do speakers take into account common ground?’, *Cognition*, vol.59, no.1, pp.91–117.
- Isaacs, E.A. and Clark, H.H. (1987) ‘References in conversation between experts and novices’, *Journal of Experimental Psychology: General*, vol.116, no.1, pp.26–37.
- Keysar, B., Barr, D.J., Balin, J.A. and Paek, T.S. (1998) ‘Definite reference and mutual knowledge: process models of common ground in comprehension’, *Journal of Memory and Language*, vol.39, no.1, pp.1–20.
- Kintsch, W. (1988) ‘The role of knowledge in discourse comprehension: a construction integration model’, *Psychological Review*, vol.95, no.2, pp.163–82.
- Krauss, R.M. and Weinheimer, S. (1967) ‘Concurrent feedback, confirmation and the encoding of referents in verbal communications; effects of referent similarity and communication mode on verbal encoding’, *Journal of Personality and Social Psychology*, vol.4, no.3, pp.343–6.
- Lakoff, G. (1987) *Women, Fire and Dangerous Things*, Chicago, University of Chicago Press.
- Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*, Cambridge, MIT Press.
- Levelt, W.J.M. and Kelter, S. (1982) ‘Surface form and memory in question answering’, *Cognitive Psychology*, vol.14, no.1, pp.78–106.
- Levin, I.P. and Gaeth, G.J. (1988) ‘How consumers are affected by the framing of attribute information before and after consuming the product’, *Journal of Consumer Research*, vol.15, no.3, pp.374–8.

- McKoon, G. and Ratcliff, R. (1992) 'Inference during reading', *Psychological Review*, vol.99, no.3, pp.440–66.
- Motley, M.T., Camden, C.T. and Baars, B.T. (1982) 'Covert formulation of anomalies in speech production – evidence from experimentally elicited slips of the tongue', *Journal of Verbal Learning and Behaviour*, vol.21, no.5, pp.578–94.
- Pickering, M.J. and Garrod, S. (in press) 'Toward a mechanistic psychology of dialogue', *Behavioral and Brain Sciences*.
- Pinker, S. (1994) *The Language Instinct*, New York, Harper-Collins.
- Sanford, A.J. and Garrod, S. (1981) *Understanding Written Language*, Chichester, John Wiley & Sons.
- Sanford, A.J. and Garrod, S. (1998) 'The role of scenario mapping in text comprehension', *Discourse Processes*, vol.26, nos2–3, pp.159–90.
- Sanford, A.J. and Sturt, P. (2002) 'Depth of processing in language comprehension: not noticing the evidence', *Trends in Cognitive Sciences*, vol.6, no.9, pp.382–6.
- Sanford, A.J., Fay, N., Stewart, A.J., Moxey, L.M. (2002) 'Perspective in statements of quantity, with implications for consumer psychology', *Psychological Science*, vol.13, no.2, pp.130–4.
- Schegloff, E.A. and Sacks, H. (1973) 'Opening up closings', *Semiotica*, vol.8, pp.289–327.
- Schober, M.F. and Clark, H.H. (1989) 'Understanding by addressees and over-hearers', *Cognitive Psychology*, vol.21, no.2, pp.211–32.
- Searle, J.R. (1980) 'Minds, brains and programs', *Behavioral and Brain Sciences*, vol.3, pp.417–24.
- Traxler, M.J. and Pickering, M.J. (1996) 'Plausibility and the processing of unbounded dependencies: an eye-tracking study', *Journal of Memory and Language*, vol.35, no.3, pp.454–75.
- Van Oostendorp, H. and De Mul, S. (1990) 'Moses beats Adam: a semantic relatedness effect on a semantic illusion', *Acta Psychologica*, vol.74, no.1, pp.35–46.
- Wilkes-Gibbs, D. and Clark, H.H. (1992) 'Coordinating beliefs in conversation', *Journal of Memory and Language*, vol.31, pp.183–94.
- Wilshire, C.E. (1999) 'The "tongue twister" paradigm as a technique for studying phonological encoding', *Language and Speech*, vol.42, no.2, pp.1–126.

---

# PART 3

# MEMORY

---

Introduction

**Chapter 8 Long-term memory: encoding to retrieval**

*Andrew Rutherford*

**Chapter 9 Working memory**

*Graham J. Hitch*

# Introduction

In Part 3 you will find two chapters dedicated to the topic of memory. Of course, you will have noticed that all the previous chapters have already included explicit references to and implicit assumptions regarding memory processes and/or memory stores. In the chapters of Part 1, the activation and utilization of stored knowledge was frequently invoked in trying to comprehend the processes of attention, perception and recognition. Similarly in Part 2, stored information (e.g. the mental lexicon) was seen to be essential to understanding categorization, language understanding and the construction of successful discourse. The fact is that memory of one sort or another is integral to every form of cognition. However, the chapters in Part 3, and also Chapter 14 in Part 5, differ from the other chapters in that they take memory as their focus of interest rather than as an important incidental to some other major topic.

In Chapter 8, ‘Long-term memory: encoding to retrieval’, the concern is to understand how information gets into and is withdrawn from memory. More than that, the emphasis is on understanding how different types of encoding and retrieval operations determine what gets remembered and in what form. The quality of memory, it turns out, results from interactions between encoding processes, the kinds of cognitive representations that are constructed, and types of retrieval operations that act upon those representations in fulfilling whatever goals a person is intent upon. One theme of the chapter is the sheer difficulty of knowing how best to conceptualize memory. A major distinction is seen between the noun *memory* and the verbs *memorising* and *remembering* or *recollecting*. That is, on the one hand, memory can be conceived as a set of stores and, on the other, memory can be thought of as a set of systems or processes. As you will see there are arguments and data that favour and count against both conceptualizations. Whichever one opts for, there is then a problem of deciding how many stores or how many processes to postulate.

One reason these questions can be so hard to answer is introduced at the start of Chapter 8. It is that the functions of memory in normal everyday cognition are so vast and diverse, and for the most part so reliable and smooth running, that – as with the processes of vision - they are really quite hard to think about. It is perhaps on account of this that one theme running throughout the chapter involves the importance of neuropsychological observations and studies for understanding the cognitive psychology of memory. Of course, it is the case that any memory impairment will itself be open to a variety of interpretations. Despite this, however, you will see in Chapter 8 that neuropsychological data have played an important part in the development of theories about the nature of memory.

In Chapter 9, ‘Working memory’, the focus of interest narrows further to take in just the memory stores and/or processes involved simply in maintaining whatever information an individual has in mind, or in executing whatever tasks they are engaged upon at a particular moment. Although there are necessarily considerable areas of overlap between the two chapters, the altered focus of interest results in a definite difference in emphasis. Where Chapter 8 dealt with issues concerning how information comes to be stored and retrieved (or not), the emphasis in Chapter 9 is more on the way in which available information is made use of. You will soon find,

however, that while the focus of interest in Chapter 9 may seem rather narrower, working memory turns out to be an extensive topic in its own right.

As the chapter explains, the notion of working memory elaborates and extends upon the older and simpler idea of a short-term memory store. Working memory is conceived as a workspace with a limited capacity. But just as in Chapter 8 it proved necessary to postulate a variety of different kinds of memory, so it turns out that working memory itself fractionates into a number of component parts. Evidence for these separate components comes from studies employing various techniques for selectively interfering with cognitive performance. Once again neuropsychological data bear strongly upon the issues, and evidence is also adduced from studies employing neuroimaging techniques.

The history of the idea of working memory provides a good illustration of a point discussed in Chapter 1. You will see how the range of application of the theory of working memory has been extended as the theory has developed, and how with this extension researchers have become more confident of their theory. Chapter 9 also provides a discussion of the importance of computer modelling in the development and testing of cognitive theories, and introduces some illustrative examples. This chapter, therefore, previews a topic that is further expanded upon in Chapters 16 and 17.

One final theme to be found in Chapter 9 is that of individual differences. As described in Chapter 1, cognitive psychology as a whole tends to play down individual differences in favour of an emphasis on what it is that people have cognitively in common. This is similar to the way in which anatomists emphasize the considerable similarities in people's bodies ahead of their individual variations. But psychology, to an even greater degree than anatomy, cannot afford to overlook individuality for long. In Chapter 9 you will see how cognitive psychologists can make use of individual differences to test their theories, and also utilize their theories to explain individual differences in cognition.



# Long-term memory: Chapter 8 encoding to retrieval

Andrew Rutherford

---

## 1 Introduction

Everyone appreciates how useful it is to have a good memory. However, fewer people appreciate that having a good memory is not just useful – it is vital to the way we live our lives and it is vital to our psychological functioning. Quite literally, our memory contains all that we know. Yet, despite the vast amount of information stored, memory almost always provides accurate and rapid access to the pertinent information we require. It is memory that tells us who we are and what we have done, it is memory that provides us with the words and grammar required to construct comprehensible sentences and it is memory that holds the information that lets us recognize different types of cars, dogs, or sporting events, or make a cup of tea or coffee. Given the essential role of memory in our lives, it is not surprising that memory has been an active area of research in psychology since its first scientific investigation by the German philosopher Hermann Ebbinghaus in the 1880s.

This chapter focuses on long-term memory, particularly episodic memory, although there will be some mention of semantic memory too. As the name suggests, **episodic memory** is a record of the episodes that constitute our lives. Episodic memory provides a description of what you have experienced (and thought) over the days, weeks and years of your life. This chapter presents some of the accounts of how episodic memory operates and some pertinent experimental evidence. Researchers interested in normal memory usually examine people with normal memories, but they also may examine people with abnormal memory resulting from physical damage to the brain. Examining the memory operation of people with brain damage may seem a peculiar way of finding out about normal memory, but an accurate account of normal memory operation also should be able to explain why and how its manner of operation changes when damage is sustained. Just as a car mechanic's understanding of the normal operation of a car engine will explain why a particular engine is not running properly, so an accurate account of normal memory should explain abnormal memory operation. More formally, it can be said that data from neuropsychological studies provide useful constraints on psychological accounts of normal memory. Of course, such studies also provide beneficial insight into the memory problems experienced by brain-damaged people.

Memory may be regarded as involving three logical stages, **encoding**, **storage** and **retrieval** (getting information in, keeping it there and then getting it back out). Typically, psychologists examine memory by presenting material and then, later, observing what can be remembered. Different manipulations can be applied at the encoding, storage and retrieval stages, depending on the purpose of the study. Investigation of any particular stage is a matter of theoretical emphasis

and experimental method, but irrespective of whether encoding, storage or retrieval is of interest, all stages will have been involved when information is remembered.

## Summary of Section 1

---

- Our long-term memory contains all that we know and all that makes us who we are.
- Usually our memory operation is very efficient.
- Episodic memory is the record of our life experiences.
- Neuropsychological findings can constrain psychological accounts of normal memory.
- Memory involves three logical stages: encoding, storage and retrieval.
- Examination of any particular stage is a matter of theoretical emphasis and experimental method.

## 2 Encoding

Encoding is the label given to the way in which objects and events in the world come to be represented in memory. Our normal perception of objects and events requires considerable encoding. However, the application of further encoding processes can produce memory representations of objects and events that differ considerably from those arising solely from perceptual processes.

### 2.1 Levels of processing

An article by Craik and Lockhart (1972) had a huge influence on memory research. At the time, the major theoretical vehicle for explaining memory performance was Atkinson and Shiffrin's (1968) '**multi-store**' or '**modal**' memory model. The 'multi-store' label referred to the assumption of separate sensory registers for each sense modality, a short-term store and a long-term memory store. (This description of different memory stores in which different memory processes operate has much in common with the multiple memory systems perspective discussed in Section 3.) The 'modal' label was due to the model encapsulating most accounts of the memory data collected up to that time (Murdock, 1967). Nevertheless, then and soon after, a number of problems were identified with the multi-store model (see Baddeley, 1997). Craik and Lockhart reviewed these problems and argued that the major determinant of the memorability of an item was not the store in which the item was held, as proposed by the multi-store model, but the level of processing that it received at encoding. Craik and Lockhart presumed that processing proceeded through a fixed sequence of levels, from early perceptual processes, through pattern recognition to the extraction of meaning. The greater the depth of processing applied to an item – the more likely it was to be remembered (see Box 8.1). Craik and Lockhart considered that, although a 'spread of elaborative coding' provided a good description of processing at the semantic level, they referred to '**depth of**

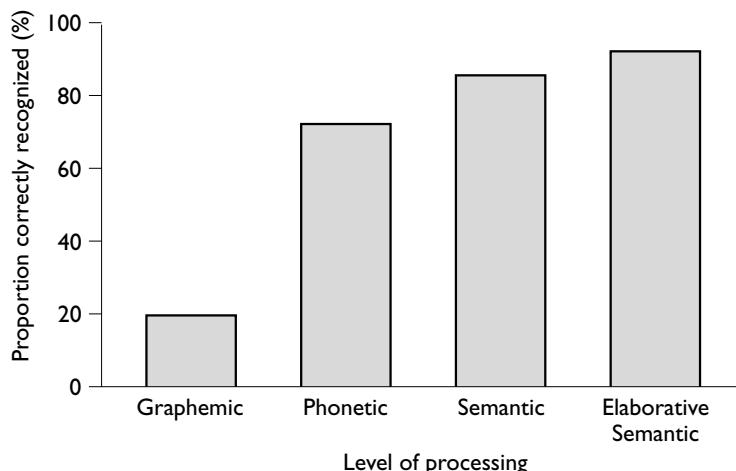
**8.1****Research study****Levels of processing**

Craik and Tulving (1975, experiment 1) reported an experiment that manipulated participants' level of processing and tested recognition memory. Participants were presented with a question followed by a word. They had to answer 'yes' or 'no' to the question and then, later on, their memory for the words was tested (see Table 8.1).

Table 8.1

<b>Question</b>	<b>Yes</b>	<b>No</b>
1 Is the word in capital letters?	TABLE	table
2 Does the word rhyme with WEIGHT?	crate	market
3 Is the word a type of fish?	shark	heaven
4 Does the word fit in the sentence? "the man peeled the _____"	orange	roof

Perceptually oriented processing must be engaged to provide answers to questions 1 and 2: graphemic for question 1 and phonetic for question 2. As the words were presented visually, visual processes always were engaged. Graphemic processing alone was engaged by question 1. To answer question 2, however, phonetic processing also must be engaged. Therefore, greater levels of processing were required to answer question 2. Questions 3 and 4 both required deeper levels of semantically oriented processing, but still more elaborative semantic processing was required to answer question 4 than question 3. The proportion of words correctly recognized as a function of the level of processing engaged at encoding is presented below.



**Figure 8.1** Recognition as a function of level of processing from Craik and Tulving, 1975

As predicted, participants' recognition memory performance increased with deeper levels of processing.

**processing**' to convey the essence of their argument. As 'deeper' levels of processing are implemented, more elaborate, longer lasting and stronger memory traces are produced. In Craik and Lockhart's conception, the processing operations both modify and leave a trace in the system. Rather than there being items that are constructed specially to be stored in memory, memories (i.e. memory traces) are simply the after-effects of processing.

Of course, processing need not proceed through all levels. The processing of information may stop at any point due to attention being diverted elsewhere or, at any given level, the processing already engaged may simply repeat rather than proceed through further levels. A common example of this sort of repetitive processing is verbally rehearsing a telephone number to keep it 'in mind' before calling the number. Craik and Lockhart labelled this **Type I processing** and considered it to manifest Primary Memory, as had been described by James (1890). **Type II processing** was the label applied to processing that proceeded through further levels. Craik and Lockhart also assumed that while Type II processing would benefit memory, no further benefit to long-term memory would accrue from repetitive Type I processing beyond that bestowed initially by the form of processing engaged.

The levels of processing framework changed the nature of psychological accounts of memory. Prior to Craik and Lockhart's (1972) article, most accounts of memory emphasized the nature of the structures holding the information to explain memory performance. Subsequently, however, most accounts of memory have emphasized the processes or mental operations carried out with respect to the material presented to explain memory performance. In the early to mid seventies, the emphasis on processing also was supported by seminal developments at the intersection of a number of cognate disciplines, such as artificial intelligence, linguistics, philosophy and neuroscience. This area of intersection is now called cognitive science and adopts a strong computational (i.e. formal processing) perspective. Nevertheless, despite all of the benefits and advantages of the levels of processing framework, it was never intended as the perfect account of memory. Objectively defining which processing levels were 'deeper' than others (and in what circumstances) was found to pose a substantial problem (Baddeley, 1978). A lack of an objective definition of levels of processing means that processing level may end up being defined in a circular fashion. Specifically, deeper levels of processing are predicted to improve memory performance, but without an objective definition of what constitutes deeper levels, improved memory performance is taken to indicate a deeper level of processing. A problem with defining processing level in this circular fashion is that the levels of processing framework predictions cannot be tested properly, as any lack of memory performance improvement can be interpreted as indicating a failure to deepen the level of processing at encoding.

Also the levels of processing framework does not provide explanations of all memory phenomena. For example, independently, Glenberg *et al.* (1977) and Rundus (1977) developed the same technique to examine Type I processing (maintenance rehearsal). Numbers were presented for participants to remember, but to stop them rehearsing the numbers, they had words to rehearse for various

intervals. However, rather than being asked to recall the numbers at test, the participants were asked to free recall the words they had been led to believe were irrelevant. In these circumstances, participants should be expected only to maintenance rehearse the words. As predicted by Craik and Lockhart's levels of processing account, it was found that the length of time spent maintenance rehearsing the words had no effect on memory as measured by free recall (Rundus, 1977). Glenberg *et al.* (1977) observed the same with free recall, but they also found that maintenance rehearsal improved recognition memory. Levels of processing can give no account of the benefit recognition memory obtains from maintenance rehearsal, not least because the levels of processing framework focuses on encoding operations and not retrieval operations. Later in the chapter we shall see how models of memory have developed to provide an account of the findings of Glenberg *et al.* (1977) and Rundus (1977).

## 2.2 Relational and item-specific processing

Psychologists have long been aware that distinctive items are well remembered (e.g. Koffka, 1935). The levels of processing framework considered that a more unique or distinctive memory trace resulted from greater depth of processing and semantic elaboration (e.g. Lockhart *et al.*, 1976). However, there is also a large body of research in psychology indicating that memory benefits from organizing items at encoding – categorizing or arranging them on the basis of properties they share (e.g. Elio and Reutener, 1970; Deese, 1959; Tulving, 1962) (see Box 8.2). These findings create something of a paradox. Establishing items' distinctiveness emphasizes their differences, while organizing items emphasizes their similarities. As Hunt and McDaniel (1993) ask, 'how can both similarity and difference be beneficial to memory?'

### 8.2

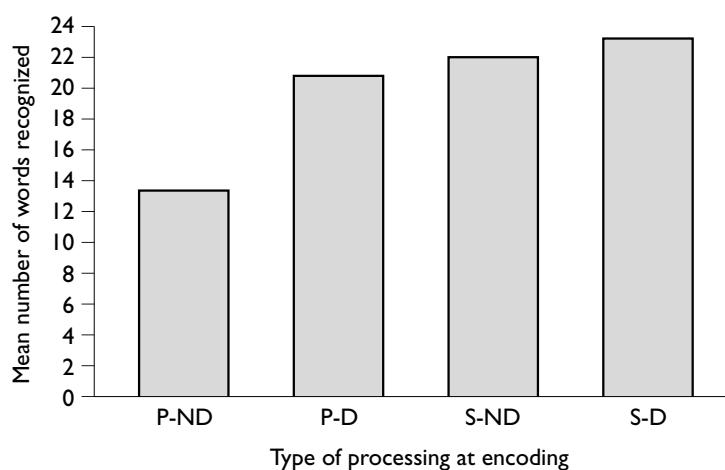
### Research study

#### Distinctive processing benefits memory independently of the level of processing

Eysenck and Eysenck (1980) conducted an experiment where distinctive processing was manipulated independently of level of processing. (Distinctive processing focuses on unique aspects of the stimulus item.)

Participants were presented with nouns that they had to process in a semantically distinct (S-D) fashion by providing a descriptor (for example, an adjective) that would be used infrequently to modify the noun. Semantically non-distinct processing (S-ND) was fostered by having participants provide a descriptor that was used frequently to modify the noun. Phonetically distinct processing (P-D) was achieved by presenting participants with nouns that are pronounced differently to the way their spelling suggests, but participants had to pronounce the words in line with their spelling. For example, the usually silent 'b' in comb would have to be pronounced at the end of the word. Phonetically non-distinct processing (P-ND) was obtained by having subjects say nouns which have conventional spelling and pronunciation.

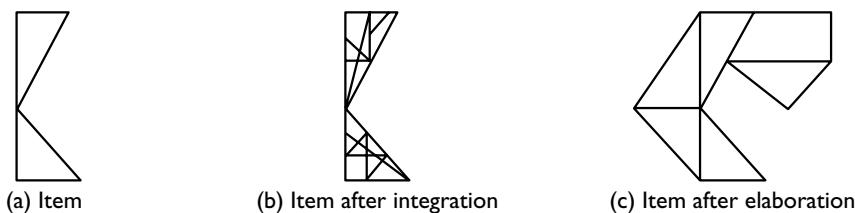




**Figure 8.2** Correct recognition as a function of experimental conditions

The results showed there was very little difference between recognition performance after semantic and distinctive, semantic and non-distinctive, and phonetic and distinctive processing, but there was a significant drop in recognition performance after phonetic and non-distinctive processing. Therefore, semantic processing enhances memory performance, but distinctive processing, even with phonetic processing, can lift memory performance to the level observed with semantic processing. In other words, it seems that distinctive processing can benefit memory performance independently of the level or depth of processing engaged.

Hunt and McDaniel (1993) resolve this paradox by referring to the different forms of processing underlying the detection of similarity and difference. **Relational processing** underlies similarity, whereas **item-specific processing** underlies distinctiveness. Mandler (1979) provides a useful description and illustration (see Figure 8.3) of the way in which memory representations are affected (organized in Mandler's terminology) by these two forms of processing. Item-specific processes focus specifically on the item's mental representation, enhancing the operation and coherence of the cognitive processes that carry the mental representation. Mandler calls this sort of enhancement 'integration'. Practising saying a word provides one example of item-specific processing, the consequence of which is greater fluency of pronunciation. In fact, enhancing the operation and coherence of cognitive processes (their integration) often is expressed as an increase in processing fluency. Relational processes establish connections between different entity representations. Mandler refers to this as 'elaboration'. Seeing a cat and thinking of it being chased by a dog is a simple example of relational processing – a relation (chasing) is drawn between two entities (cat and dog). According to Mandler (1979), maintenance rehearsal results in integration (i.e. employs item-specific processing), while semantic processing results in elaboration (i.e. employs relational processing).



**Figure 8.3** A graphic analogy of integration (due to item-specific processing) and elaboration (due to relational processing) (Mandler, 1979). Item-specific processing (b) enhances the coherence of the cognitive processes carrying the mental representation components (depicted by the links between the components of the representation). Relational processing (c) establishes connections between the mental representations of the target item and other items

### 2.2.1 Encoding processing and Mandler's (1980) dual process model of recognition

Soon after his account of integration and elaboration in memory representations, Mandler presented a very influential dual-process model of recognition (Mandler, 1980). In this model, one process runs very quickly and is based on familiarity. The sense of familiarity is thought to result from processing fluency, that is, the more fluently an item can be processed (or encoded) the more familiar it feels. Familiarity depends upon the degree of integration of the entity representation: greater integration makes the presented item feel more familiar and facilitates subsequent processing of the same or similar items (Jacoby and Dallas, 1981). The other process runs more slowly and employs more involved and extensive search and retrieval operations used in recall to determine if the entity was presented before. The search and retrieval process benefits from elaboration – the greater the elaboration, the greater the benefit to the retrieval process. Presumably, the connections between representations established by relational processing provide a variety of different routes (cues) to the representation of the target. (See Box 8.3 for a more detailed discussion of item-specific relational processing.)

Mandler also distinguishes between simple recognition and identification. Simple recognition is based only upon an evaluation of the familiarity of an entity and, therefore, provides a context-free judgement of prior occurrence. In contrast, identification employs a search and retrieval stage, as well as a familiarity evaluation. Search and retrieval processes first provide and then employ contextual information. This is used to restrict the memory search on subsequent retrieval cycles. For example, if someone is trying to remember a person's name, usually they know (i.e. can retrieve) if it is a male or female name, and they may even know (or guess) the place where they frequently encounter this person. Both gender and place provide contexts that are able to restrict or focus the memory search. Mandler also assumes that both familiarity and search and retrieval processes are initiated simultaneously and operate in parallel. However, as the speedy familiarity-based process will finish first, time-pressured recognition is most likely based on simple recognition.

**8.3****Research study**

### **Effects of item-specific and relational processing on free recall and recognition**

Hunt and Einstein (1981, experiment 1) presented participants with either a categorized list of 36 words (6 words from each of 6 categories) or 36 unrelated words. It was assumed that participants would process the categorized words spontaneously in a relational fashion (but not necessarily in an item-specific fashion), while participants receiving the unrelated words would process them spontaneously in an item-specific fashion (but not necessarily in a relational fashion).

For both categorized and unrelated lists, free recall and recognition were tested. However, prior to these tests, participants were required either to sort the words into specified categories (a relational processing task), or to rate the pleasantness of the words (an item-specific processing task). Participants read a short story for one minute before trying to free recall the 36 words. Recognition was tested after free recall. Table 8.2 below presents the average free recall and recognition scores.

Table 8.2

	Categorized list		Unrelated list	
	Relational processing	Item-specific processing	Relational processing	Item-specific processing
Free recall <sup>1</sup>	.42	.48	.47	.33
Recognition <sup>2</sup>	.73	.93	.89	.91

<sup>1</sup>Correct free recall as a proportion of total number of items presented (i.e. 36).

<sup>2</sup>AG scores – a nonparametric measure of recognition sensitivity (Pollack *et al.*, 1964).

Free recall of the categorized list was greater after item-specific processing (.48) than after relational processing (.42), but free recall of the unrelated list was greater after relational processing (.47) than after item-specific processing (.33). Therefore, free recall benefits from task processing that is different from that facilitated by the type of list. This shows that both relational and item-specific processing contribute to free recall.

Although considerable research has shown that recognition memory benefits from relational processing (e.g. Craik and Tulving, 1975), Hunt and Einstein's recognition data do not simply replicate the free recall data. Recognition of categorized list items was greater after item-specific processing (.93) than after relational processing (.73) but, unlike free recall, recognition of unrelated list items was the same irrespective of relational (.89) or item-specific processing (.91). It seems additional item-specific processing may become redundant for free recall, but it continues to benefit recognition.

Mandler's (1979, 1980) descriptions provide an explanation for the findings obtained by Glenberg *et al.* (1977) and Rundus (1977). Free recall derives greatest benefit from relational processing, but little benefit from maintenance rehearsal (i.e.

item-specific processing), which promotes integration. In contrast, familiarity-based simple recognition depends upon the degree of integration. Therefore, a high degree of item-specific processing, maintenance rehearsal or Type I processing, will benefit recognition to a greater degree than it will benefit free recall.

To explain the effects on memory performance of different forms of encoding requires consideration of the relations between memory encoding, memory representation and memory tests. This illustrates the point made in Section 1: whether interest is in encoding, storage or retrieval, all stages of memory are involved when information is remembered.

## Summary of Section 2

---

- The levels of processing framework was presented as a counter to the multi-store memory model.
- The levels of processing framework asserted that memorability was due to the level of processing received at encoding and not the store in which the item was held.
- Distinctive processing can benefit memory independently of the level of processing.
- Relational and item-specific are two important types of processing.
- Mandler's dual-process model of recognition memory assumes item-specific processing enhances processing fluency or familiarity, as well as the distinctiveness, of an item, while relational processing supports context-based retrieval.
- Recall derives greater benefit from relational processing, while recognition derives greater benefit from item-specific processing.

## 3 Memory stores and systems

A memory store is where non-active memory representations are held. For example, imagine your favourite item of clothing. When not in use, the memory representation upon which this image depends will be held in a memory store. Memory systems include memory stores, but memory systems also include all the processes that operate when memory representations are active, such as the processes that generate the image of your favourite item of clothing. The memory systems perspective is that memory stores and memory processing are localized in the same part of the brain. This view receives support from research on connectionist systems, where representation and processing are intimately related. The accounts to be presented in this section certainly assume that memory and its associated processing are localized within the brain. However, as will be described, these accounts have tended to focus on identifying different types of memory systems and their apparent locations in the brain, rather than on describing the processing and nature of the memory representation.

### 3.1 Multiple memory systems

Tulving and associates (e.g. Tulving and Schacter, 1990; Schacter *et al.*, 2001) are strong advocates of a multiple memory systems perspective. (Table 8.3 presents the various systems and subsystems of human learning and memory proposed by Schacter and Tulving, 1994.) Although Schacter and Tulving present five long-term memory (LTM) systems and eleven sub-systems, discussion here will concentrate on the distinction between episodic and semantic memory.

**Episodic memory** is considered to be a record of a person's experiences. It stores information about the events and occurrences that make up a person's life and, crucially, according to Wheeler *et al.* (1997), the subjective experiences that accompany the information retrieved from episodic memory. Therefore, the answers to questions such as, 'What did you do yesterday afternoon?' and 'Have you seen this picture before?' would tax episodic memory. **Semantic memory** is considered to be our general knowledge store. In short, it contains all the information underlying our understanding of the world. For example, it provides the information we use to recognize or describe different types of animals, objects, etc., it provides the information for using and understanding language and it stores the sort of information we would employ to choose our ideal summer holiday destination. Questions such as 'What is the capital of Scotland?' and 'Did Plato own a car?' would tax semantic memory. However, no personal experience accompanies the information retrieved from semantic memory.

**Table 8.3** Schacter and Tulving's (1994) systems and subsystems of human learning and memory.

System	Other labels	Subsystems	Retrieval type
Procedural	Non-declarative	(i) Motor skills (ii) Cognitive skills (iii) Simple conditioning (iv) Simple associative learning	Implicit
Perceptual representation	Non-declarative	(i) Visual word form (ii) Auditory word form (iii) Structural description	Implicit
Semantic	General Factual Knowledge	(i) Spatial (ii) Relational	Implicit
Primary	Working	(i) Visual (ii) Auditory	Explicit
Episodic	Personal Autobiographical Events		Explicit

In terms of research focus, a distinction certainly exists between episodic and semantic memory. Most psychology texts identify Collins and Quillian's (1969) research as seminal work on the topic of semantic memory. They converted a system for representing information in computer systems into a model of human knowledge and examined its psychological reality. This and further investigation established the study of human knowledge, or semantic memory, as a distinct research area with its own issues, paradigms and measures.

One criticism of Tulving's distinction between episodic and semantic memory systems is the need for substantial communication between them. This is illustrated by the fact that information encoded in episodic memory usually is comprehended fully, yet our knowledge of the world, upon which this comprehension is based, would be stored in semantic memory. To provide episodic memory with easy access to semantic memory information, Tulving (1984) suggested episodic memory was embedded within semantic memory. A study reported by Anderson and Ross (1980) is relevant to this issue. They investigated the independence of semantic and episodic memory systems, and were interested in whether episodic memory information affected semantic memory. Two types of task can be used to examine semantic and episodic memory. A sentence *verification* task requires participants to state whether a sentence is true or false and is regarded as a test of semantic memory. A sentence *recognition* task requires participants to state whether or not a sentence was presented earlier and is regarded as a test of episodic memory. Anderson and Ross measured how long participants took to verify a sentence. For example: a spaniel is a dog. (Here 'dog' is the category and 'spaniel' is an exemplar of that category, cf. Chapter 5.) Beforehand, participants were allocated to one of five conditions. In four of these conditions, participants were presented with episodic information about the categories and exemplars. This information was presented in the form of simple sentences that participants had to learn (for example: a plumber pets a dog, a spaniel retrieves a ball). In the fifth control condition, participants received no information about the category or the exemplar. The results revealed that the time taken to verify sentences (that is, to make semantic judgements) was affected by the nature of the episodic information about the exemplar and category presented in the previous sentences. Contrary to there being a distinct separation between episodic and semantic memory, episodic information affected retrieval from semantic memory.

The need to facilitate transfer of information from the semantic memory system to the episodic memory system led Tulving (1984) to suggest episodic memory was embedded within semantic memory, while the results of the Anderson and Ross study reveal that information also transfers from the episodic memory system to the semantic memory system. Such transfer between systems raises the question, why should there be separate episodic and semantic memory systems? Anderson and Ross note that semantic memory must respond to experience, but the manner in which this occurs is not specified. This last point is related to another criticism of distinct episodic and semantic memory systems dealt with below.

The multiple memory systems perspective, especially the episodic and semantic memory distinction, has been criticized as lacking theoretical development (e.g. McKoon *et al.*, 1986; Neely, 1989). In particular, the way in which different variables differentially affect the operation of episodic and semantic memory

systems has not been described. For example, Anderson (1974) demonstrated the fan effect. The fan effect is the name given to the phenomenon where participants' recognition times for sentences about a particular concept increase as more information about the concept is acquired (see Anderson, 2000). As a recognition task is employed, it is episodic memory that is tested and, indeed, the fan effect is observed in tests of episodic memory but not in tests of semantic memory (Shoben *et al.*, 1978). McKoon *et al.* (1986) point out that although these observations could be presented as support for a distinction between episodic and semantic memory systems, the theoretical account of these systems provides no basis for predicting the fan effect in episodic rather than semantic memory tests. Indeed, the completely opposite result (i.e. detecting the fan effect in semantic, but not in episodic memory tests) also could be presented as support for the distinction between semantic and episodic memory systems. A model cannot be specified sufficiently when both of two contradictory patterns of effects can be interpreted as supporting the model.

Rather than develop the theory underlying the proposed multiple memory systems, so that unambiguous theoretical predictions can be made, the tendency has been simply to categorize memory systems and sub-systems by identifying them with particular types of memory performance. However, another criticism of multiple memory systems is the lack of agreement, even among multiple memory systems proponents, on the criteria by which systems and sub-systems are distinguished and classified. For example, Johnson and Chalfonte (1994) consider episodic and semantic memory to be two sub-systems rather than two separate systems. Yet another criticism is that a lack of agreement on the criteria by which systems are distinguished and classified may lead to a spurious proliferation of systems (e.g. Roediger *et al.*, 1999).

Frequently, neuroimaging techniques are used to identify the brain regions associated with performance on these different tasks and memory tests. These brain regions have been interpreted, somewhat simply, as the neuroanatomical sites of the particular memory systems underlying the different tasks and tests. Recently, however, there has been an increase in the application of more sophisticated neuroanatomical network analysis approaches. These examine the interactions between different memory 'systems' underlying performance on different tasks and memory tests (e.g. Nyberg and Cabeza, 2001). As discussed in Section 5.2.2, interactions between systems raise interesting questions about what constitutes a system.

Neuropsychological data obtained from the study of amnesic patients (see Box 8.4) also have been presented to support the distinction between episodic and semantic memory. Tulving (e.g. Tulving, 1983) argues that the amnesic syndrome is due to a severe deficit in episodic memory combined with an intact semantic memory. The retention of amnesics' intellect and language skills is strong evidence that a substantial part of their semantic memory operates normally. However, the apparently normal operation of semantic memory appears to arise from the use of semantic information acquired prior to the amnesic trauma. Gabrieli *et al.* (1988) noted that HM (see Box 8.4) continued to use many of the verbal expressions common at the time of his operation in the 1950s, and was only mildly successful in explaining words and phrases that had come into use since then. Even after

considerable practice learning the meaning of ten unfamiliar words, HM was exceedingly poor at matching the words to their definitions. Grossman (1987) reported similar problems in amnesic patients suffering from Korsakoff's Syndrome (in which patients have damage to their brains in similar areas to HM), while Cermak and O'Connor (1983) describe how an amnesic patient, who had been a laser expert, was able to explain new developments in laser technology after reading a recent article. However, a little later, he could not remember anything of what he had read and could not provide answers to questions based on what he had read. Contrary to Tulving and Schacter's claims, therefore, the amnesic syndrome cannot be attributed to a severe deficit in just episodic memory. As deficits are observed across both semantic and episodic memory tasks, the nature of the amnesic syndrome does not support a distinction between independent episodic and semantic memory systems.

## 8.4

## Research study

### The Amnesic syndrome

Milner (1966) described the case of HM. In 1953, when he was 27, HM underwent brain surgery in an attempt to treat intractable epilepsy. The aim was to remove those parts of his brain considered to be the focus of the epileptic seizures. The operation was a success in that subsequently, the epilepsy could be controlled by drugs, but a tragic and unforeseen result of the operation was that HM became profoundly amnesic. The removal of the anterior two thirds of the hippocampus from both sides of the brain (bilaterally) is thought to have been responsible for his amnesia (e.g. Squire, 1987). Although HM retained his memory for events occurring up to a short time before the operation, he seemed to have lost most of his ability to form new memories. HM stayed with his parents for some time after the operation. However, as HM's memory problems make it impossible for him to live without supervision, he has lived in a nursing home since 1980. HM's father died in 1967 and his mother died ten years later. Yet, six years after moving to the nursing home, HM thought he still lived with his mother and was unsure if his father was alive (Parkin, 1993). HM can read the same book or magazine repeatedly without any recollection of having done so before and, typically, after spending all morning with psychologists doing various tests, he cannot remember the testing session, nor recognize the psychologists when they return in the afternoon.

As even this brief account might suggest, despite his substantial memory impairment, and in common with other amnesics, HM is able to interact and converse quite normally. He also retains a normal immediate memory span and demonstrates memory for a variety of perceptual and motor tasks, although he reports no memory of the learning episodes.

The amnesic syndrome seems to manifest whenever there is bilateral hippocampal damage. Although there may be a variety of different reasons for such damage, Korsakoff Syndrome provides the largest group. Korsakoff Syndrome is caused by a thiamine deficiency, often associated with chronic alcoholism, which leads to damage to parts of the brain, including the hippocampus.

As we have seen, a distinction between episodic and semantic memory is a very useful heuristic for distinguishing between types of memory task and research areas, but it is unlikely that Tulving's descriptions of separate episodic and semantic memory systems is correct. A simpler conception is that semantic memory is an abstraction of episodic experience. Common aspects of episodes are, by definition, experienced repeatedly. In contrast, there is an inconsistent association between the common aspects of the episodes and the various contexts in which they occur. As a result, the common aspects of the episodes will be well learned and will be able to be retrieved easily and speedily, while the associated contexts, without the benefit of such repetition, will become inaccessible or will fade from memory (e.g. Baddeley, 2002; Hintzman, 1986). An account almost identical to this, based on connectionist memory research, has been presented by McClelland *et al.* (1995).

### 3.2 Declarative and procedural memory

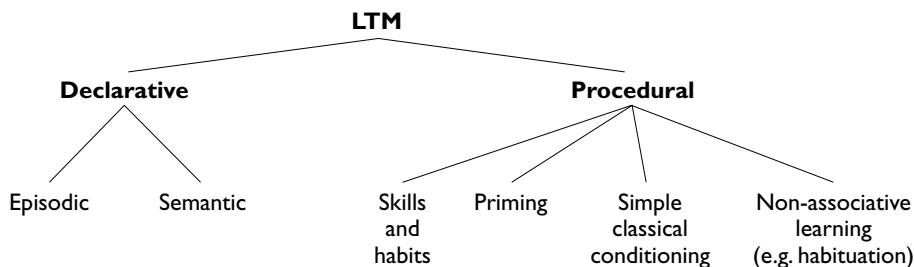
One influential systems account of the amnesic syndrome was presented by Squire (e.g. Cohen and Squire, 1980). Squire proposes two separate LTM systems: a *declarative* system and a *procedural* system (see Figure 8.4). The declarative–procedural distinction was made with respect to knowledge by the philosopher Ryle (1949) and is much used in cognitive science (e.g. Winograd, 1975).

**Declarative knowledge** corresponds to ‘knowing that’. Responses to semantic and episodic memory tasks typically provide declarative information, such as ‘(I know that) the capital of Scotland is Edinburgh’, or ‘(I know that) I have seen that picture before’. Cohen (1984) described declarative knowledge as being represented in a system ‘... in which information is ... first processed or encoded, then stored in some explicitly accessible form for later use, and then ultimately retrieved upon demand’.

**Procedural knowledge** corresponds to ‘knowing how’. For example, the type of information underlying the ability to ride a bicycle is procedural knowledge. Cohen (1984) describes procedural knowledge as being involved when ‘experience serves to influence the organization of processes that guide performance without access to the knowledge that underlies the performance’. One way to access this information is to observe performance of a procedure that employs the information: try riding a bike and observe what you do and when, and consider why you do it.

It has been known for some time that amnesics are able to exhibit normal or close to normal learning on a variety of different tasks. For example, the time HM takes to complete a jigsaw puzzle declines with practice. Squire organizes the tasks on which amnesics demonstrate learning under the headings of skills and habits, priming, simple classical conditioning and non-associative learning (see Figure 8.4). Although amnesics' performance on these sorts of tasks demonstrates that learning occurs, typically, amnesics cannot remember having carried out any of the tasks before.

According to Squire, it is a failure of the declarative memory system that produces the deficits observed in amnesic memory performance (for example, the



**Figure 8.4** Squire's LTM distinctions and their relation to LTM tasks. Declarative memory involves conscious remembrance of events and facts. Procedural memory encompasses a variety of different abilities where experience alters behaviour without there being conscious access to the memory content.

Source: adapted from Squire, 1992

inability to remember having practised the task), while the continued operation of the procedural memory system explains the learning amnesics are able to exhibit. However, procedural memory is considered to be 'a heterogeneous collection of separate abilities that can be additionally dissociated from each other' (Squire *et al.*, 1993). A number of different processes or memory systems seem necessary to serve the variety of different tasks labelled as examples of procedural memory. Achieving an understanding of the different components underlying procedural memory is an important contemporary goal (Baddeley, 1997).

## Summary of Section 3

---

- Non-active memory representations are held in a memory store.
- The memory systems perspective regards memory storage and processing as occurring within a system that is localized within the brain.
- The multiple memory systems perspective advocates a large number of memory systems and sub-systems, including episodic and semantic memory.
- Evidence from normal participants and amnesics, as well as theoretical concerns, argues against the multiple memory systems perspective, particularly regarding episodic and semantic memory.
- Semantic memory may develop from abstracted episodic memory information.
- The less elaborate distinction between procedural and declarative memory systems may provide a more accurate account of long-term memory.
- It is likely that procedural memory fractionates into a number of different memory systems.

## 4 Retrieval

Retrieval is the label given to the way in which information held in memory is made available for use. Retrieval involves finding, activating and sometimes further processing pertinent memory representations.

### 4.1 Encoding specificity and transfer appropriate processing

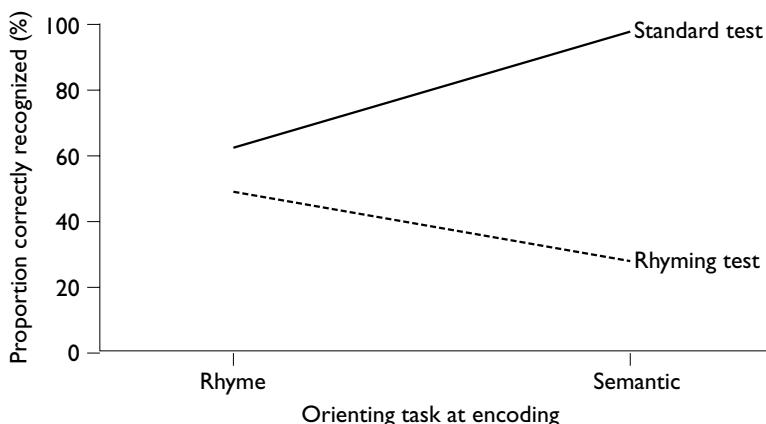
The notions of **encoding specificity** (e.g. Tulving, 1983) and **transfer appropriate processing** (e.g. Bransford *et al.*, 1979) continue to influence research and accounts of memory retrieval. The encoding specificity hypothesis was introduced by Tulving and Osler (1968) in relation to a study of the role of cues in memory retrieval. They presented participants with target words written in capitals. Also presented with each target word were zero, one, or two weakly associated words written in lower case (for example, MUTTON, fat, leg: CITY, dirty, village). Participants were told that the words in lower case might help them remember the capitalized target words and to try and think about how the lower case words were related to the target words. Tulving and Osler found a single weak associate aided recall of the target word, provided the weak associate had been presented at learning. Neither one nor two weak associates aided recall if they had not been presented at learning – recall was not assisted by the provision of these cues at test alone. Tulving and Osler concluded that specific retrieval cues facilitate recall only if information about them and their relation to the target item is stored along with the target item. Successful retrieval of the target item increases with the overlap between the information stored in memory and the information employed at retrieval (Tulving, 1979).

The transfer appropriate processing (TAP) account also emphasizes the overlap between encoding and retrieval. Morris *et al.* (1977) presented TAP as an adjunct to Craik and Lockhart's (1972) levels of processing framework to give proper emphasis to retrieval processing, which they believed had been neglected. Therefore, TAP focuses on the overlap between the *processes* engaged at encoding and the *processes* engaged at retrieval. Specifically, it predicts that the best memory performance will be observed when the processes engaged at encoding transfer appropriately to retrieval (see Box 8.5).

Although encoding specificity deals with information and TAP deals with processing, these distinctions may be different sides of the same coin. Both accounts emphasize the relationship between encoding and retrieval, and the benefit to memory performance when encoding conditions are recapitulated at retrieval. As information at encoding and at retrieval is manifest within the cognitive system by psychological processes, it may be more a matter of expression rather than psychological substance whether information or processes are recapitulated.

**8.5****Research study****An experimental test of transfer appropriate processing**

Morris *et al.* (1977) conducted an experiment to test the TAP hypothesis. All participants were presented with a list of words, such as CAT and TABLE. For half of the participants the orienting questions were of the form, Does the word rhyme with hat? Does the word rhyme with label? (phonetic processing), while the other participants received questions of the form, Is it an animal? Do you sit at it? (semantic processing). The next day, half of the participants in the phonetic orienting condition were given a standard, semantically oriented recognition test (for example, identify which of the following words were presented previously: CAT, ROAD, POUND, TABLE, BALL, and so on), while the other half were shown another set of words and asked to identify (that is, recognize) which words rhymed with the words presented the day before (for example, identify which of the following words rhyme with those presented previously: FIRE, MAT, STAIR, CABLE, PAPER, etc.). Similarly, half of the semantic orienting condition participants received a standard, semantically oriented recognition test, while the others received the rhyme test. Figure 8.5 below presents the mean proportion of correctly recognized words as a function of orienting and recognition tasks.



**Figure 8.5** Correct recognition as a function of encoding task and type of test

When rhyme/phonetic processing was employed at encoding and test, memory performance was better than when semantic processing was employed at encoding but rhyme/phonetic processing was employed at test. Likewise, when semantic processing was employed at encoding and test, memory performance was better than when rhyme/phonetic processing was employed at encoding and semantic processing was employed at test. As TAP predicts, memory performance was better when there was a match between the processes engaged at encoding and the processes engaged at retrieval.

## Summary of Section 4

---

- Retrieval involves finding, activating and sometimes further processing pertinent memory representations.
- Both encoding specificity and TAP emphasize the relationship between encoding and retrieval, such that performance is enhanced by increasing similarity of information (encoding specificity) or processing (TAP).

## 5 Implicit memory

Free recall, cued recall and recognition memory tests are explicit tests of memory. When any of these techniques are employed, it is clear to participants that their memory is being tested – it is plain that memory must be used to do the task. However, it is also possible to test participants' memory without them appreciating that their memory is being used. When this is done, the memory test is said to be implicit. This terminology follows Roediger *et al.* (1992), who define the learning task as either incidental or intentional and the memory test as either explicit or implicit. Unfortunately, however, the terms applied in this research area have been varied and mixed. For example, Schacter and Tulving (1994) refer to both task and test as being explicit or implicit, Jacoby (1984) refers to the test as being incidental or intentional, while both Johnson and Hasher (1987) and Richardson-Klavehn and Bjork (1988) refer to the test as being direct or indirect and label the type of memory taxed as being explicit or implicit. Just to make things a little more complicated, there is also an area of research labelled *implicit learning*. This is concerned with the way in which rule-governed relations between stimulus items are learned without conscious awareness. Although it seems that work in implicit learning should have consequence for implicit memory, these two research areas remain quite separate. In the following sections, only implicit memory research will be considered and the terminology of Roediger *et al.* (1992) will be employed.

### 5.1 Perceptual and conceptual implicit memory

Roediger and McDermott (1993) list a variety of tests used to investigate *perceptual* and *conceptual* incidental memory. The word-fragment task employed in the Tulving *et al.* (1982) study (see Box 8.6) is an example of a perceptual implicit memory test. Perceptual (or data-driven) implicit tests require participants to resolve perceptually impoverished displays (McDermott and Roediger, 1996). A display is perceptually impoverished if it presents a version of the stimulus that is not as easily identified as is usual, due to the relatively poor quality of the stimulus, the short duration of the stimulus presentation, or to the stimulus presented being incomplete. To identify the stimulus, it is assumed that processes involving the analysis of perceptual or surface-level features are engaged, although other representations needed for stimulus identification also may be involved (Mulligan, 1998). In addition to word-fragment completion, other tests of perceptual implicit memory include word-stem completion, where a whole word has to be completed from only the first few letters (e.g. Graf *et al.*, 1984); word (perceptual) identification, where

participants have to identify words presented very swiftly (for example, for 35 milliseconds, Jacoby and Dallas, 1981); anagram solution (e.g. Srinivas and Roediger, 1990); and lexical decision (e.g. Duchek and Neely, 1989). In contrast, conceptual implicit tests require participants to employ their semantic knowledge to answer questions or provide responses to a cue (McDermott and Roediger, 1996) and so they are assumed to engage processes that involve the analysis of semantic information (Mulligan, 1998). Although less research has been carried out on conceptual implicit memory, a number of tests have been developed. They include, word association (Shimamura and Squire, 1984) and category instance generation – where participants have to generate examples of a particular category (e.g. Srinivas and Roediger, 1990) and answering general knowledge questions (Blaxton, 1989). Irrespective of whether the tests are perceptual or conceptual, implicit memory is demonstrated when better performance occurs with recently presented items compared with items not presented recently.

## 8.6

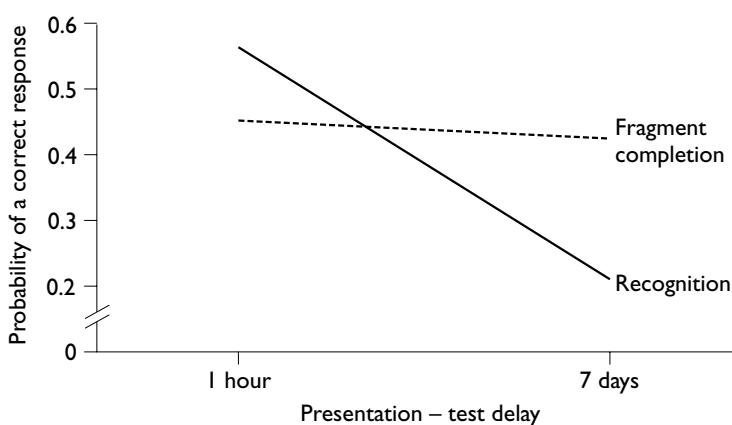
## Research study

### Empirical evidence of implicit memory

Tulving *et al.* (1982) conducted an experiment in which participants were asked to try to learn a list of 96 words. One hour later, the participants were asked to carry out a recognition test that used 24 of the presented words (targets) and 24 similar words that had not been presented before (distractors), and a word-fragment completion test, where 24 word fragments were based on another set of 24 presented words and 24 word fragments were based on words not presented before. (Word-fragment completion involves the presentation of real words with certain letters removed. For example, the word-fragment F\_O\_ \_ A\_L might be presented and participants would complete the fragment by replacing the empty slots with O, T, B and L to provide the real word, FOOTBALL. In this study, each word fragment had only one real word solution. For word fragments based on presented words, the presented words were the only real word solutions). Seven days later, participants received recognition and word-fragment completion tests, as described above, for the remaining 48 words presented originally.

Participants were expected to carry out the word-fragment task without realizing that half of the solutions to the word-fragments are words that they had been presented with before. On this basis, it is assumed that the word-fragment task is an implicit test of memory. The interesting measure for the word-fragment test is how many word fragments were completed correctly when the corresponding full word had been presented previously compared with the number of correct completions when a corresponding full word had not been presented previously. Tulving *et al.* found more word fragments were completed when the corresponding full word had been presented previously and labelled this a word repetition priming effect. The figure below presents the probability of correct response as a function of type of test (word-fragment or recognition) after one hour and after 7 days.





**Figure 8.6** Probability of a correct response as a function of the presentation-test delay and type of test

While explicit memory performance on the recognition test declined substantially between the test that occurred one hour after and the test that occurred 7 days after the stimulus presentation, there was no significant decline in implicit memory performance between the word-fragment completion tests at one hour and at 7 days after stimulus presentation.

The word-fragment priming effect could occur because participants realize the words shown originally also complete the word fragments. Participants then might try to recall words and try to match them to the word fragments, converting the implicit test into an explicit test. However, if this occurred, then performance on the word-fragment completion test after 7 days should have declined in line with the explicit recognition test performance.

## 5.2 Accounts of implicit memory

The distinction between implicit and explicit memory tasks provides a *description* of a person's psychological experience of memory use as different tasks are done (Schacter, 1987). It does not provide an *explanation* of the different effects observed with perceptual implicit memory tests, conceptual implicit memory tests and explicit memory tests. So far, most research has focused on transfer appropriate processing or memory systems accounts to explain these phenomena. These accounts will be outlined and considered in turn, but it soon will be appreciated that neither of these accounts is able to accommodate all of the research findings. Nevertheless, as research continues it is important to know the strengths and weaknesses of previous accounts, so new theoretical formulations may retain the former and avoid the latter.

### 5.2.1 TAP account

Roediger and associates (e.g. Roediger *et al.*, 1989) have been the strongest advocates for applying Morris *et al.*'s (1977) TAP account to explain the differences in performance on implicit and explicit memory tests. According to Roediger

and associates, the important distinction is not between implicit and explicit retrieval from different memory stores, but the match between the type of (perceptual or conceptual) processing engaged when stimuli are tested. Roediger and associates argued that in most experiments on implicit memory, the processing required at memory test often was confounded with the implicit-explicit memory test distinction. TAP predicts that perceptual implicit tests will benefit most from encoding that engaged similar perceptual processes. Likewise, performance on conceptual implicit memory tests will benefit from encoding that engaged similar conceptual processes. Therefore, TAP predictions regarding performance on conceptual implicit memory tests are identical to TAP predictions for explicit memory tests. Indeed, Roediger and Blaxton (1987) state that performance on all implicit conceptual tests should match that observed with free recall, as free recall is the definitive conceptual test. In free recall, no retrieval cues are provided, so participants must rely exclusively on top-down conceptual processing.

However, there are research findings at odds with the TAP account. For example, Hunt *et al.* (1990) noted that orthographic distinctiveness (a perceptual factor) affected both (perceptual) implicit memory test performance and free recall. McDermott and Roediger (1996) also report that while presenting words that were conceptually related to each target word (conceptual repetition) enhanced the free recall of the target words, it did not enhance performance in the category exemplar generation test. Similarly, conceptual repetition by virtue of a picture followed by a corresponding word (or vice versa) also enhanced free recall, but again had no effect on priming in the category exemplar generation test. McDermott and Roediger did obtain enhanced priming in the category exemplar generation test after verbal conceptual repetition when participants were given relational processing instructions. However, the difference between participants' performance on the implicit conceptual memory test (category exemplar generation) and on the explicit memory test (free recall) contradicts the TAP prediction of equivalent (conceptual processing based) memory performance. Therefore, TAP is able to give a good account of much, but not all, of the implicit and explicit memory test data.

In an attempt to deal with these problems, Roediger and associates modified the TAP account and relabelled it *components of processing* (e.g. Roediger *et al.*, 1999). Essentially, this view considers performance on different memory tests to involve different sets of processes. The sets of processes employed by different memory tests may share some processes (i.e. component processes), but different processing components will be employed in any two tests that dissociate. Roediger's TAP account of implicit memory has been very influential in focusing research on the nature of the processing underlying encoding at learning and retrieval when memory is tested. However, the TAP account has been criticized for being circular. For example, the TAP account states that repetition priming occurs when there is appropriate transfer of processing, but, unfortunately, the mark of appropriate transfer of processing is considered to be repetition priming. (As mentioned in Section 2.1, Baddeley (1978) criticized levels of processing for a similar circularity of account.) Greater detail on the mechanisms operating in these

circumstances is necessary to avoid this circularity and indeed, this is one of the requirements placed on the components of processing account by McDermott and Roediger (1996).

### 5.2.2 Memory systems accounts

Both Tulving and associates' multiple memory systems perspective and Squire's declarative and procedural memory systems have been applied to give account of the differences observed between explicit and implicit memory test performance. In both cases, the differences in explicit and implicit memory test performance are regarded as being due to the tests taxing different memory systems (see Table 8.3 and Figure 8.4). Squire simply attributes performance on explicit memory tasks to the declarative memory system and performance on implicit memory tasks to the procedural memory system. As procedural memory is very likely to fractionate into a number of different memory systems, there is greater similarity between Squire's account and the multiple memory systems perspective than may appear at first glance. The multiple memory systems perspective attributes performance on perceptual implicit memory tasks such as word priming and fragment completion to the visual word form subsystem of the perceptual representation system, while picture priming is attributed to the structural description subsystem of the perceptual representation system (e.g. Schacter *et al.*, 2001). Also within this perspective, Gabrieli (1999) attributes performance on conceptual implicit memory tests to yet another system – the conceptual representation system (this compares with the perceptual representation system, see Table 8.3). Schacter (1990) attempted to shed some light on the operation of the perceptual representation system by suggesting that it operates according to TAP principles.

In Section 3.1, one of the criticisms of the multiple memory systems perspective was that a lack of agreement on the criteria by which systems are distinguished and classified may lead to a spurious proliferation of systems. In fact, as more and more memory systems are postulated, so the difference between a processing perspective and the multiple memory systems perspective diminishes. A 'system' has to be more than just the brain structures that carry out the cognitive operations for a specific task. As, ultimately, all cognitive processes have a neural basis, simply defining a memory system as the brain structures that carry out the cognitive operations for a specific task goes no further than stating where in the brain these processes run. Identifying where a process runs does not distinguish between the processing perspective and the multiple memory systems perspective (e.g. Crowder, 1993). Similarly, as neuroanatomical network analysis reveals that the brain structures involved in memory are highly interactive, rather than being stand-alone systems (e.g. Nyberg and Cabeza, 2001), so the difference between the multiple memory systems and processing perspectives diminishes.

## 5.3 Implicit memory and amnesia

While amnesics perform poorly on explicit memory tests, their performance on implicit memory tests is similar to that of controls. For example, Graf *et al.* (1984) presented lists of words to amnesics and controls who had to judge how much they liked each word. Later, participants received four memory tests: three explicit (free

recall, cued recall, recognition) and one implicit (word-stem completion). As expected, amnesics performed much more poorly on the explicit memory tests than controls, but they exhibited as much implicit memory as controls on the word-stem completion test.

Vaidya *et al.* (1995) found no difference between amnesics and controls in either perceptual implicit memory performance (word-fragment completion) or conceptual implicit memory (word association). However, amnesics' performance on explicit perceptual and conceptual memory tests was as poor as expected. Similar findings were reported by Cermak *et al.* (1995).

These results present problems for the TAP account of implicit memory. According to the TAP account, the reason amnesics are able to perform implicit memory tests on a par with normal controls is because they retain their perceptual processing capability. Therefore, amnesics' poor memory performance should be due to impaired conceptual processing. However, the ability of amnesics to perform conceptual implicit memory tests on a par with normal controls contradicts this account. Moreover, the fact that amnesics exhibited their usual poor memory performance on explicit perceptual and conceptual memory tests indicates that the distinction between implicit and explicit memory tests is more important than the distinction between perceptual and conceptual processing.

Cermak *et al.* (1995) explained their findings in terms of dual memory processes, such as underlie Mandler's account of recognition outlined earlier. According to Cermak *et al.*, amnesics will exhibit normal memory performance whenever the memory task can be accomplished on the basis of item familiarity-processing fluency. Usually, implicit memory tasks can be accomplished on this basis, whereas explicit tasks usually require more context-based discriminations. Likewise, perceptual tasks often can be accomplished on the basis of item familiarity, while conceptual tasks typically require context-based discrimination processing. However, both familiarity and context-based processing may be applied to any task. Of course, the exact nature of the task will determine how successfully it can be accomplished using familiarity or context-based processing. It is the varying degrees of success in applying familiarity or context-based processing to a task that give rise to the differences between some implicit and explicit memory tasks, and between some perceptual and conceptual processing tasks.

## Summary of Section 5

---

- An explicit memory task taxes memory with participants' awareness, but an implicit memory task taxes memory without participants' awareness.
- Free recall, cued recall and recognition are standard explicit memory tasks.
- There are conceptual and perceptual implicit memory tasks.
- Perceptual implicit memory tasks include: word-fragment completion, word-stem completion, word identification, anagram solution and lexical decision.

- Conceptual implicit tests include: word association, category instance generation and answering general knowledge questions.
- Amnesics exhibit normal memory performance on implicit tasks.
- The multiple memory systems perspective and Squire's declarative and procedural memory systems attribute the differences between explicit and implicit memory test performance to these tasks being served primarily by different memory systems.
- The TAP account attributes the differences between explicit and implicit memory test performance to perceptual implicit tests benefiting most from perceptual encoding at presentation, while conceptual implicit memory tests and standard explicit memory tests benefit from conceptual encoding at presentation.
- Cermak *et al.* suggest implicit memory tasks can be accomplished on the basis of item familiarity/processing fluency, whereas explicit tasks usually require more context-based discriminations.

## 6 Jacoby's process-dissociation framework

Although some tasks and memory tests are regarded as providing good measures of certain encoding and retrieval processes, it would be wrong to think they provide pure measures of these processes. Irrespective of the task and memory test employed, it is likely that the specific memory processes under investigation will be contaminated to some extent by the operation of other memory processes. This point is especially relevant with respect to implicit and explicit memory performance.

It was mentioned in Box 8.6 that participants might convert the implicit test into an explicit test if they realized that many of the word fragments (or word stems or anagrams) corresponded with words shown earlier. One approach to this issue was presented by Jacoby (e.g. Jacoby, 1991), who assumes that implicit memory performance is based primarily on automatic (familiarity-based) processes (see the discussion of Mandler's dual-process model in Section 2.2.1), while explicit memory depends most on conscious recollective memory processes. Box 8.7 outlines Jacoby's process-dissociation procedure and also shows how the measures of automatic and recollective processes derived from the procedure not only confirm theoretical expectations, but also provide some insight into the mechanisms underlying memory effects.

Although Jacoby's inventive approach and its developments (e.g. Yonelinas, 2002) offer new and attractive methods for understanding and investigating memory, the validity of Jacoby's assumption that recollective and automatic processes are independent has provoked considerable debate and research. Joordens and Merikle (1993) claim that only automatic processes retrieve items from memory. Recollective processes only operate to acquire further information about these words. As only automatic processes are involved in the retrieval of items from memory, recollective processes do not contribute to memory

retrieval *per se*. According to Joordens and Merikle, therefore, with respect to memory retrieval, rather than recollective processes being independent of automatic processes, recollective processes are redundant in relation to automatic processes. Jacoby (e.g. Jacoby *et al.*, 1997) strongly disputes this claim and has provided a description of the conditions necessary for the implementation of a tenable process-dissociation procedure (Jacoby, 1998). Research continues on this and other issues, such as whether all automatic familiarity-based retrieval is unconscious and whether all controlled recollective retrieval is conscious (e.g. Gardiner *et al.*, 1998).

## 8.7

## Research study

### Process-disassociation procedure

Jacoby *et al.* (1993) presented words to participants under a full attention condition, where they just read the words, and under a divided attention condition, where they also had to listen to a tape-recorded list of numbers and indicate each time a sequence of three odd numbers was presented. The aim of the divided attention task was to reduce the influence of recollective processes at memory test, but to leave automatic processes unaffected. Later, participants received a word-stem completion memory test where half of the word stems were coloured green and half were red. When presented with a green word stem, participants had to use it as a cue to remember one of the words presented earlier. If they could not remember a word, they were asked to complete the word stem with the first word that came to mind. When presented with a red word stem, participants again were asked to use it as a cue to remember one of the words presented earlier, but they were not to provide this as a response – instead they were to complete the stem to make some other word that came to mind. The green stem task is an inclusion test and the red stem task is an exclusion test (see below). Jacoby *et al.* found that the probabilities of responding with a previously presented word were as follows:

Probability of responding with a previously presented word		
Attention	Inclusion test	Exclusion test
Full	0.61	0.36
Divided	0.46	0.46

On an inclusion test, the probability of responding with a presented word equals the probability of conscious recollection ( $R$ ), plus the probability that this word is remembered automatically ( $A$ ) when there is a failure of conscious recollection ( $1 - R$ ). However, remembering the word automatically, given a failure of conscious recollection, is a conditional probability that is obtained by multiplying the probability of automatic remembering and the probability of a failure of conscious



recollection. Therefore, the probability of responding with a presented word on an inclusion test is:

Equation one

$$\text{Inclusion} = R + A(1 - R)$$

On an exclusion test, the probability of providing a presented word equals the probability of remembering automatically when there is a failure of conscious recollection. Therefore, the probability of providing a presented word on an exclusion test is:

Equation two

$$\text{Exclusion} = A(1 - R)$$

Equations one and two may be rewritten to obtain the probabilities of conscious recollection ( $R$ ) and of remembering automatically ( $A$ ). That is:

$$R = \text{Inclusion} - \text{Exclusion}$$

$$A = \frac{\text{Exclusion}}{(1 - R)}$$

$R$  and  $A$  estimates for the words presented in the second part of the experiment, based on the data presented above are as follows:

Attention	R	A
Full	0.25	0.47
Divided	0.00	0.46

These estimates are consistent with the view that automatic memory processes are unaffected by changes in the attentional resource available at encoding, whereas recollective processes suffer severely if focused attentional resources are not deployed at encoding. Nevertheless, the calculation of  $R$  as zero should be interpreted only as indicating that participants' recollective component may have been insufficient to register under these particular experimental conditions (Baddeley, 1997).

## Summary of Section 6

- Jacoby's process-dissociation framework assumes that two independent processes contribute to memory performance: automatic and recollective memory processes.
- Automatic (familiarity-based) processes are assumed to be unconscious.
- Recollective (search-and-retrieval-based) memory processes are assumed to be under conscious control.
- Implicit memory performance is based primarily on automatic processes.
- Explicit memory performance is based primarily on recollective memory processes.

## 7 Remember and know judgements

Tulving (1985) carried out the first experiment requiring a distinction to be made between items recalled due to *remembering* that the item was presented (you have a conscious recollection of the item appearing in the study) and *knowing* the item was presented (you simply know that the item appeared but you have no conscious recollection of its occurrence). According to Tulving, remembering should reflect retrieval from episodic memory, while knowing should reflect retrieval from semantic memory. In his typical neologistic fashion, Tulving created and applied the label *autonoetic* (self-knowing) to the form of consciousness accompanying retrieval from episodic memory and the label *noetic* (knowing) to the form of consciousness accompanying retrieval from semantic memory (see Box 8.8).

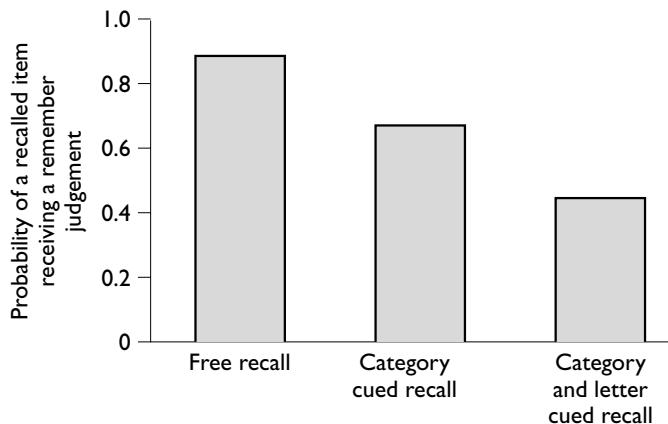
Tulving's (1985) study employed free recall and cued recall, but most other studies of remember and know judgements have focused on recognition for two reasons. First, there was an initial presumption that remember and know judgements were relevant to dual-process accounts of recognition (see Section 2.2). Second, while both recall and recognition tests provide a good proportion of remember judgements, only recognition tests provide a good proportion of know judgements – few know judgements are obtained with recall.

The subjective nature of remember and know judgements should be highlighted. In a memory experiment employing recall, participants provide remember and know judgements only after they have recalled an item. When recognition is employed, a one-step or two-step procedure can be applied. With one-step procedures, participants straight away judge whether they remember, know or were not presented with an item. All items judged as remember or know are deemed to be recognized. With two-step procedures, remember or know judgements are made only after the participant positively recognizes an item. (Know judgements seem to be more accurate when a two-step procedure is used, Eldridge *et al.*, 2002). As the experimenter knows which words have been presented, an objective decision can be made about the accuracy of the recalled or recognized item. However, remember and know judgements cannot be assessed objectively, as they are based on the extent to which participants *believe* their introspections concord with the remember and know descriptions provided. Remember and know judgements are employed because they provide information on states of awareness that it seems impossible to obtain from more conventional, objective measures. For example, experimental groups may obtain identical recognition scores, but they may differ in terms of their proportions of remember and know judgements (Gardiner and Richardson-Klavehn, 2001). To improve the accuracy of remember and know judgements, Gardiner (e.g. Gardiner *et al.*, 1998) suggests that participants should be provided with the opportunity to indicate that the recalled or recognized item was a guess. Without this facility, guesses will be placed in the know category by default, so affecting the validity of the remember – know procedure.

**8.8****Research study****Empirical evidence of a distinction between remembering and knowing**

Tulving (1985) reported two experiments. In experiment I, participants studied pairs of words. The first word of the pair specified a category and this was followed by an exemplar of that category (for example, fruit – PEAR). Three memory tests then were presented: free recall of each exemplar (in any order), cued recall with the category name as the cue, and cued recall with the category name and the first letter of the exemplar as the cue. In all tests, participants had to judge whether their responses were accompanied by a feeling of *remembering* or a feeling of *knowing*. Item recall was scored in a particular fashion: all of the items free recalled were scored, but with the category name cued recall test, only items not free recalled were scored, and with the category name and first letter cued recall test, only items not free recalled nor recalled on the basis of category name cues were scored.

Tulving reasoned that items free recalled had the richest representation in episodic memory as they had been recalled without any cues. Items recalled only on the basis of category name cueing had a less rich representation in episodic memory because they required cueing. Items recalled only on the basis of category name and first letter cueing had the poorest representation in episodic memory because they required most cueing. As feelings of remembering (indicated by remember judgements) arise as a consequence of the representational richness of episodic memory, remember judgements should be most prevalent with free recall items, less prevalent with category name cued recall items and least prevalent with category name plus first exemplar letter cued recall items. Data analysis revealed that the probability of a recalled item receiving a remember judgement was a function of the type of memory test, just as Tulving had predicted.



**Figure 8.7** Probability of a recalled item receiving a remember judgement as a function of the recall test

In experiment 2, Tulving presented participants with the same tasks, but while testing on half of the stimulus items occurred immediately, the other stimulus items were tested after eight days. Compared with immediate testing, the probability of a remember judgement decreased after an eight day presentation-test gap. All of these findings are consistent with Tulving's view that remember judgements reflect the rich information available in episodic memory, which diminishes over longer retention intervals.

Gardiner (2002) identifies four types of variable in terms of their effect on remember and know judgements. There are variables that increase the number of remember responses, but do not affect know responses (for example, levels of processing, Gardiner, 1988). There are variables that increase know responses, but do not affect remember responses (for example, suppression of focal attention during stimulus presentation prior to test, Mantyla and Raudsepp, 1996). There are variables that increase know responses and decrease remember responses (nonword versus word presentation, Gardiner and Java, 1990). Finally, there are variables that have similar effects on remember and know responses (for example, long and short response deadlines, Gardiner *et al.*, 1998). Gardiner claims that as some variables exert similar effects on remember and know responses, while other variables exert different effects on remember and know responses, distinct memory processes must underlie know and remember responses.

## 7.1 Do remember and know judgements reflect different response criteria?

Donaldson (1996) argued that remember and know judgements simply reflect decisions based on different response criteria. Rather than reflecting qualitatively different memory processes, Donaldson's detection theory account attributes remember and know judgements to different criterial points on a single quantitative dimension of memory strength. Gardiner *et al.* (2002) have presented considerable evidence contradicting this account. However, the focus here will be on a different strand of contradictory evidence.

According to Donaldson's detection model, the strongest memories are associated with remember judgements. Yet, as Gardiner and Conway (1999) point out, 'knowing' is the natural state accompanying answers to semantic memory questions – conscious recollection of the encoding event(s) very rarely accompanies the retrieval of information from semantic memory. Does this mean that semantic memory information has less strength than episodic type information? A good indication of the answer to this question was provided by a large-scale naturalistic study conducted by Conway *et al.*, (1997). They examined changes in awareness as psychology knowledge was acquired by undergraduates. Psychology students took a three-alternative multiple-choice test (MCT) and six months later, they took the same test again. For each question, the MCT correct answer involved information presented directly in a lecture, while the plausible but incorrect MCT answers involved information also presented in the same lecture. The students had to select one of the MCT answers and then indicate whether they (i) remembered a learning

episode where they encountered this information, ii) just knew that this was the correct answer, that is, they had a strong feeling of knowing but did not remember a learning episode, iii) neither remembered the learning episode or knew the answer but felt the chosen answer was more familiar, or iv) felt they were guessing, for example, choosing the example that looked least unlikely. (The familiarity category was included to separate aspects of the know judgement, but it has no bearing on the results discussed here.) Table 8.4 presents the response probabilities for correct answers over the two tests.

**Table 8.4** Response probabilities for correct answers over the two multiple choice tests

Response	Probability of correct answers in test 1	Probability of correct answers in test 2
remember	.39	.14
know	.19	.43
familiar	.25	.26
guess	.17	.17

Source: Table 1 in Conway *et al.*, 1997

Over the two tests, the proportion of familiar (iii) and guess (iv) judgements remained the same. However, there was an interesting pattern of change for the proportion of remember and know judgements over the two tests. In Test 1, remember judgements dominated, with a low proportion of know judgements. However, six months later, in Test 2, know judgements dominated, with a low proportion of remember judgements. There is a substantial ‘remember to know’ shift in the proportion of judgements made about correct answers over the six month gap between tests. This finding applied to all of the students participating in the study, but the shift from remember to know judgements was most pronounced for students who attained the highest grades.

Contrary to Donaldson’s detection theory account, these data indicate that know (and not remember) judgements are associated with the stronger type of memory (the information most likely to be remembered). Conway *et al.* (1997) interpret the ‘remember-to-know’ shift as revealing the way in which memories are modified by the loss of detail, so that a more abstract version is retained as conceptual knowledge in semantic memory. These data and their interpretation are consistent with the view that semantic memory is an abstraction of episodic memory regularities and contrast with Tulving’s (1984) conception that the episodic memory system is embedded within the semantic memory system (see Section 3.1).

## Summary of Section 7

---

- Remembered items may be given a remember judgement (you have a conscious recollection of the item appearing in the study) or a know judgement (you simply know the item appeared but you have no conscious recollection of its occurrence).

- Remember and know judgements are based on the extent to which participants believe their introspections concord with remember and know descriptions – they are subjective judgements.
- Contrary to Donaldson's detection theory account, know judgements appear to reflect stronger memories than remember judgements.

## 8 Conclusions

Craik and Lockhart's levels of processing article stimulated a great deal of research on memory encoding processes. This work emphasized that the mental operations carried out on presented material had great consequence for the memorability of this material. However, work by Mandler, Tulving, and Morris, Bransford and Franks demonstrates that good memory performance relies upon the interaction between memory encoding, memory representation and retrieval operations.

Around the same time as Craik and Lockhart's levels of processing article, Tulving provided a description of separate semantic and episodic memory systems, but it was not until the 1980s that memory systems research began to exert a substantial theoretical influence. This influence seems to have arisen as a consequence of a number of somewhat related factors, including a renaissance in connectionist research and developments in cognitive neuroscience, particularly with respect to neuroimaging techniques, and cognitive neuropsychological investigation of abnormal memory as a consequence of brain damage. Tulving and associates' multiple memory systems perspective, particularly the distinction between episodic and semantic memory, was criticized heavily by cognitive psychologists, the majority of whom found greater evidence for Squire's simpler procedural/declarative distinction. The multiple memory systems perspective was more warmly received in the field of neuropsychology. Nevertheless, enthusiasm for the multiple memory systems perspective has waned for a variety of reasons. One reason is the observation that amnesics' performance on tasks that tax new semantic information and new episodic information seems to be affected equally. Another reason is the lack of development of the theoretical accounts of the various multiple memory systems. Yet another reason is the weakening of the conception of distinct memory systems, as a result of the number of memory systems proposed and the substantial system interactions identified by neuroanatomical network analysis.

In recent years, there has been a marked increase in research activity focusing on retrieval operations. Initially, this interest was prompted by two phenomena: implicit memory and remember and know judgements. Research on these topics reveals the benefit of the theoretical constraints imposed by neuropsychological findings. Meanwhile Jacoby and associates' work on the process-dissociation procedure not only provides theoretical insight into these phenomena, but also has introduced new methods to investigate memory. Due to the nature of the phenomena considered, retrieval research has had to confront and accommodate issues of consciousness, as well as the fact that people can modify and change how they retrieve information from memory. Each of these factors has contributed to an overall improvement in theoretical accounts of memory.

An aim of this chapter was to present an overview of research in the psychology of memory that not only reflects these influences and changes, but also demonstrates the exciting advances in understanding that these perspectives have provided. Memory research continues to be one of the most active research areas in psychology, where useful and interesting theoretical and methodological developments are leading to a more accurate appreciation of memory operation.

## Further reading

- Yonelinas, A.P. (2002) ‘Components of episodic memory: the contribution of recollection and familiarity’ in Baddeley, A., Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.
- Tulving, E. (2002) ‘Episodic memory and common sense: how far apart?’ in Baddeley, A., Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.
- Baddeley, A. (2002) ‘The concept of episodic memory’ in Baddeley, A., Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.
- Yu, J. and Bellezza, F.S. (2000) ‘Process dissociation as source monitoring’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.26, no.6, pp.1518–33.

## References

- Anderson, J.R. (1974) ‘Retrieval of propositional information from long-term memory’, *Cognitive Psychology*, vol.6, pp.451–74.
- Anderson, J.R. (2000) *Cognitive Psychology and its Implications*, New York, Worth Publishers.
- Anderson, J.R. and Ross, B.H. (1980) ‘Evidence against a semantic–episodic distinction’, *Journal of Experimental Psychology: Human Learning and Memory*, vol.6, pp.441–65.
- Atkinson, R.C. and Shiffrin, R. (1968) ‘Human memory: a proposed system and its control processes’ in Spence, K. and Spence, J. (eds) *The Psychology of Learning and Motivation Vol.2*, New York, Academic Press.
- Baddeley, A.D. (1978) ‘The trouble with levels: a re-examination of Craik and Lockhart’s framework for memory research’, *Psychological Review*, vol.85, pp.139–52.
- Baddeley, A.D. (1997) *Human Memory: Theory and Practice* (revised edn) Hove, Psychology Press.
- Baddeley, A.D. (2002) ‘The concept of episodic memory’ in Baddeley, A., Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.
- Blaxton, T.A. (1989) ‘Investigating dissociations among memory measures: support for a transfer-appropriate processing framework’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.15, pp.657–68.

- Bransford, J.D., Franks, J.J., Morris, C.D. and Stein, B.S. (1979) 'Some general constraints on learning and memory research' in Cermak, L.S. and Craik, F.I.M. (eds) *Levels of Processing in Human Memory*, Hillsdale, NJ, LEA.
- Cermak, L.S. and O'Connor, M. (1983) 'The anterograde and retrograde retrieval ability of a patient with amnesia due to encephalitis', *Neuropsychologia*, vol.21, pp.213–34.
- Cermak, L.S., Verfaelie, M. and Chase, K.A. (1995) 'Implicit and explicit memory in amnesia: an analysis of data-driven and conceptually driven processes', *Neuropsychology*, vol.9, pp.281–90.
- Cohen, N.J. (1984) 'Preserved learning capacity in amnesia: evidence for multiple memory systems' in L.R. Squire and N. Butters (eds) *Neuropsychology of Memory*, New York, Guildford Press.
- Cohen, N.J. and Squire, L.S. (1980) 'Preserved learning and retention of pattern-analysing skill in amnesia using perceptual learning', *Cortex*, vol.17, pp.273–8.
- Collins, A.M. and Quillian, M.R. (1969) 'Retrieval time from semantic memory', *Journal of Verbal Learning and Verbal Behavior*, vol.8, pp.240–47.
- Conway, M.A., Gardiner, J.M., Perfect, T.J., Anderson, S.J. and Cohen, G. (1997) 'Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates', *Journal of Experimental Psychology: General*, vol.126, pp.393–413.
- Craik, F.I.M. and Lockhart, R.S. (1972) 'Levels of processing: a framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, vol.11, pp.671–84.
- Craik, F.I.M. and Tulving, E. (1975) 'Depth of processing and the retention of words in episodic memory', *Journal of Experimental Psychology: General*, vol.104, pp.268–94.
- Crowder, R. (1993) 'Systems and principles in memory theory: another critique of pure memory' in Collins, A.F., Gathercole, S.E., Conway M.A. and Morris P.E. (eds) *Theories of Memory*, Hove, LEA.
- Deese, J. (1959) 'Influence of inter-item associative strength upon immediate free recall', *Psychological Reports*, vol.5, pp.305–12.
- Donaldson, W. (1996) 'The role of decision processes in remembering and knowing', *Memory and Cognition*, vol.24, pp.523–33.
- Duchek, J.M. and Neeley, J.H. (1989) 'A dissociative word frequency x levels of processing interaction in episodic recognition and lexical decision tasks', *Memory and Cognition*, vol.17, pp.148–62.
- Eldridge, L.L., Sarfatti, S. and Knowlton, B.J. (2002) 'The effect of testing procedure on remember–know judgements', *Psychonomic Bulletin and Review*, vol.9, pp.139–45.
- Elio, R.E. and Reutener, D.B. (1970) 'Colour context as a factor in encoding and as an organization device for retrieval of word lists', *Journal of General Psychology*, vol.99, pp.223–32.
- Eysenck, M.W. and Eysenck, M.C. (1980) 'Effects of processing depth, distinctiveness and word frequency on retention', *British Journal of Psychology*, vol.71, pp.263–74.

- Gabrieli, J.D.E. (1999) ‘The architecture of human memory’ in Foster, J.K. and Jelic, M. (eds) *Memory: Systems, Process or Function?* Oxford, Oxford University Press.
- Gabrieli, J.D.E., Cohen, N.J. and Corkin, S. (1988) ‘The impaired learning of semantic knowledge following bilateral medial-temporal lobe resection’, *Brain*, vol.7, pp.157–77.
- Gardiner, J.M. (1988) ‘Functional aspects of recollective experience’, *Memory and Cognition*, vol.16, pp.309–13.
- Gardiner, J.M. (2002) ‘Episodic memory and autonoetic consciousness: a first person approach’ in Baddeley, A. Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.
- Gardiner, J.M. and Conway, M.A. (1999) ‘Levels of awareness and varieties of experience’ in Challis, B.H. and Velichkovsky, B.M. (eds) *Stratification in Cognition and Consciousness*, Amsterdam, John Benjamins Publishing.
- Gardiner, J.M. and Java, R.I. (1990) ‘Recollective experience in word and nonword recognition’, *Memory and Cognition*, vol.18, pp.23–30.
- Gardiner, J.M., Ramponi, C. and Richardson-Klavehn, A. (1998) ‘Experiences of remembering, knowing and guessing’, *Consciousness and Cognition*, vol.7, pp.1–26.
- Gardiner, J.M., Ramponi, C. and Richardson-Klavehn, A. (2002) ‘Recognition memory and decision processes: a meta-analysis of remember, know and guess responses’, *Memory*, vol.10, pp.83–98.
- Gardiner, J.M. and Richardson-Klavehn, A. (2001) ‘Remembering and knowing’ in Tulving, E. and Craik F.I.M. (eds) *The Oxford Handbook of Memory*, Oxford, Oxford University Press.
- Glenberg, A.M., Smith, S.M. and Green, C. (1977) ‘Type 1 rehearsal: maintenance and more’, *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.339–52.
- Graf, P., Squire, L.R. and Mandler, G. (1984) ‘The information that amnesic patients do not forget’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.10, pp.164–78.
- Grossman, M. (1987) ‘Lexical acquisition in alcoholic Korsokoff psychosis’, *Cortex*, vol.23, pp.631–44.
- Hintzman, D.L. (1986) ““Schema abstraction” in a multiple-trace memory model”, *Psychological Review*, vol.93, pp.411–28.
- Hunt, R.R. and Einstein, G.O. (1981) ‘Relational and item-specific information in memory’, *Journal of Verbal Learning and Verbal Behavior*, vol.20, pp.497–514.
- Hunt, R.R., Humphrey, N. and Toth, J.P. (1990) ‘Perceptual identification, fragment completion and free recall: concepts and data’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.16, pp.282–90.
- Hunt, R.R. and McDaniel, M.A. (1993) ‘The enigma of organization and distinctiveness’, *Journal of Memory and Language*, vol.32, pp.421–45.
- Jacoby, L.L. (1984) ‘Incidental versus intentional retrieval: remembering and awareness as separate issues’ in Squire, L.R. and Butters, N. (eds) *Neuropsychology of Memory*, New York, Guildford Press.

- Jacoby, L.L. (1991) 'A process dissociation framework: separating automatic from intentional uses of memory', *Journal of Memory and Language*, vol.30, pp.513–41.
- Jacoby, L.L. (1998) 'Invariance in automatic influences on memory: toward a user's guide for the process dissociation procedure', *Journal of Experimental Psychology: Learning Memory and Cognition*, vol.24, pp.3–26.
- Jacoby, L.L. and Dallas, M. (1981) 'On the relationship between autobiographical memory and perceptual learning', *Journal of Experimental Psychology: General*, vol.3, pp.306–40.
- Jacoby, L.L., Toth, J.P. and Yonelinas, A.P. (1993) 'Separating conscious and unconscious influences of memory: measuring recollection', *Journal of Experimental Psychology: General*, vol.122, pp.139–54.
- Jacoby, L.L., Yonelinas, A.P. and Jennings, J.M. (1997) 'The relation between conscious and unconscious (automatic) influences: a declaration of independence' in Cohen, J.D. and Schooler, J.W. (eds) *Scientific Approaches to Consciousness*, Mahwah, NJ, LEA.
- James (1890) *The Principles of Psychology*, New York, Henry Holt and Company.
- Johnson, M.K. and Chalfonte, B.L. (1994) 'Binding complex memories: The role of reactivation and the hippocampus' in Schacter, D.L. and Tulving, E. (eds) *Memory Systems 1994*, Cambridge, MA, MIT Press.
- Johnson, M.K. and Hasher, L. (1987) 'Human learning and memory', *Annual Review of Psychology*, vol.38, pp.631–68.
- Joordens, S. and Merikle, P.M. (1993) 'Independence or redundancy? Two models of conscious and unconscious influences', *Journal of Experimental Psychology: General*, vol.122, pp.462–7.
- Koffka, K. (1935) *Principles of Gestalt Psychology*, New York, Harcourt Brace.
- Lockhart, R.S., Craik, F.I.M. and Jacoby, L.L. (1976) 'Depth of processing, recognition and recall' in Brown, J. (ed.) *Recall and Recognition*, London, Wiley.
- Mandler, G. (1979) 'Organization and repetition: organizational principles with special reference to rote learning' in Nilsson, L-G. (ed.) *Perspectives on Memory Research*, Hillsdale, NJ, LEA.
- Mandler, G. (1980) 'Recognizing: the judgement of previous occurrence', *Psychological Review*, vol.87, pp.252–71.
- Mantyla, T. and Raudsepp, J. (1996) 'Recollective experience following suppression of focal attention', *European Journal of Cognitive Psychology*, vol.8, pp.195–203.
- McClelland, J.L., McNaughton, B.L. and O'Reilly, R.C. (1995) 'Why there are complimentary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory', *Psychological Review*, vol.102, pp.419–57.
- McDermott, K.B. and Roediger, H.L. (1996) 'Exact and conceptual repetition dissociate conceptual memory tests: problems for transfer appropriate processing theory', *Canadian Journal of Experimental Psychology*, vol.50, pp.57–71.

- McKoon, G., Ratcliff, R. and Dell, G.S. (1986) 'A critical examination of the semantic/episodic distinction', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.12, pp.295–306.
- Milner, B. (1966) 'Amnesia following operation on the temporal lobes' in Whitty, C.W.M. and Zangwill, O.L. (eds) *Amnesia*, London, Butterworths.
- Morris, C.D., Bransford, J.D. and Franks, J.J. (1977) 'Levels of processing versus transfer appropriate processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.519–33.
- Mulligan, N.W. (1998) 'The role of attention during encoding in implicit and explicit memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.24, pp.27–47.
- Murdock, B.B. (1967) 'Recent developments in short-term memory', *British Journal of Psychology*, vol.58, pp.421–33.
- Neely, J.H. (1989) 'Experimental dissociations and the semantic/episodic memory distinction' in Roediger, H.L. and Craik, F.I.M. (eds) *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving*, Hillsdale, NJ, LEA.
- Nyberg, L. and Cabeza, R. (2001) 'Brain imaging of memory' in Tulving, E. and Craik F.I.M. (eds) *The Oxford Handbook of Memory*, Oxford, Oxford University Press.
- Parkin, A. (1993) *Human Memory*, Oxford, Blackwell.
- Pollack, I., Norman, D.A. and Galatner, E. (1964) 'An efficient nonparametric analysis of recognition memory', *Psychonomic Science*, vol.1, pp.327–8.
- Richardson-Klavehn, A. and Bjork, R.A. (1988) 'Measures of memory', *Annual Review of Psychology*, vol.39, pp.475–543.
- Roediger, H.L. and Blaxton, T.A. (1987) 'Retrieval modes produce dissociations in memory for surface information' in Gorfein, D. and Hoffman, R.R. (eds) *Memory and Cognitive Processes: The Ebbinghaus Centennial Conference*, Hillsdale, NJ, LEA.
- Roediger, H.L., Buckner, R.L. and McDermott, K.B. (1999) 'Components of processing' in Foster, J.K. and Jelic, M. (eds) *Memory: Systems, Process or Function?*, Oxford, Oxford University Press.
- Roediger, H.L. and McDermott, K.B. (1993) 'Implicit memory in normal human subjects' in Spinnler, H. and Boller, F. (eds) *Handbook of Neuropsychology*, Amsterdam, Elsevier.
- Roediger, H.L., Weldon, M.S. and Challis, B.H. (1989) 'Explaining dissociations between implicit and explicit measures of retention: a processing account' in Roediger, H.L. and Craik, F.I.M. (eds) *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, Hillsdale, NJ, LEA.
- Roediger, H.L., Weldon, M.S., Stadler, M.L. and Riegler, G.L. (1992) 'Direct comparison of two implicit memory tests: word fragment and word stem completion', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.18, pp.1251–69.
- Rundus, D. (1977) 'Maintenance rehearsal and single level processing', *Journal of Verbal Learning and Verbal Behaviour*, vol.16, pp.665–81.

- Ryle, G. (1949) *The Concept of Mind*, London, Hutchinson.
- Schacter, D.L. (1987) ‘Implicit memory: history and current status’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.13, pp.501–18.
- Schacter, D.L. (1990) ‘Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate’ in Diamond, A. (ed.) *The Development and Neural Bases of Higher Cognitive Functions*, New York, New York Academy of Sciences.
- Schacter, D.L. and Tulving, E. (1994) ‘What are the memory systems of 1994?’ in Schacter, D.L. and Tulving, E. (eds) *Memory Systems*, 1994, Cambridge, MA, MIT Press.
- Schacter, D.L., Wagner, A.D. and Buckner, R.L. (2001) ‘Memory systems of 1999’ in Tulving, E. and Craik F.I.M. (eds) *The Oxford Handbook of Memory*, Oxford, Oxford University Press.
- Shimamura, A.P. and Squire, L.R. (1984) ‘Paired associate learning and priming effects in amnesia: a neuropsychological approach’, *Journal of Experimental Psychology: General*, vol.113, pp.556–70.
- Shoben, E.J., Wescourt, K.T. and Smith, E.E. (1978) ‘Sentence verification, sentence recognition and the semantic-episodic distinction’, *Journal of Experimental Psychology: Human Learning and Memory*, vol.4, pp.304–17.
- Squire, L.R. (1987) *Memory and Brain*, New York, Oxford University Press.
- Squire, L.R. (1992) ‘Declarative and nondeclarative memory: multiple brain systems supporting learning and memory’, *Journal of Cognitive Neuroscience*, vol.4, pp.232–43.
- Squire, L.R., Knowlton, B. and Musen, G. (1993) ‘The structure and organization of memory’, *Annual Review of Psychology*, vol.44, pp.453–95.
- Srinivas, K. and Roediger, H.L. (1990) ‘Classifying implicit memory tests: category association and anagram solution’, *Journal of Memory and Language*, vol.29, pp.389–412.
- Tulving, E. (1962) ‘Subjective organization in free recall of “unrelated” words’, *Psychological Review*, vol.69, pp.344–54.
- Tulving, E. (1979) ‘Relation between encoding specificity and levels of processing’ in Cermak, L.S. and Craik, F.I.M. (eds) *Levels of Processing in Human Memory*, Hillsdale, NJ, LEA.
- Tulving, E. (1983) *Elements of Episodic Memory*, Oxford, Clarendon Press.
- Tulving, E. (1984) ‘Précis of Elements of episodic memory’, *Behavioral and Brain Sciences*, vol.7, pp.223–68.
- Tulving, E. (1985) ‘Memory and consciousness’, *Canadian Psychology*, vol.26, pp.1–12.
- Tulving, E. and Osler, S. (1968) ‘Effectiveness of retrieval cues in memory for words’, *Journal of Experimental Psychology*, vol.77, pp.593–601.
- Tulving, E. and Schacter, D.L. (1990) ‘Priming and human memory systems’, *Science*, vol.247, pp.301–6.

- Tulving, E., Schacter, D.L. and Stark, H.A. (1982) 'Priming effects in word-fragment completion are independent of recognition memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.8, pp.336–42.
- Vaidya, C.J., Gabrieli, J.D.E., Keane, M.M. and Monti, L.A. (1995) 'Perceptual and conceptual memory processes in global amnesia', *Neuropsychology*, vol.10, pp.529–37.
- Wheeler, M.A., Stuss, D.T. and Tulving, E. (1997) 'Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness', *Psychological Bulletin*, vol.121, pp.331–54.
- Winograd, T. (1975) 'Frame representations and the declarative procedural controversy' in Bobrow, D. and Collins, A. (eds) *Representation and Understanding: Studies in Cognitive Science*, New York, Academic Press.
- Yonelinas, A.P. (2002) 'Components of episodic memory: the contribution of recollection and familiarity' in Baddeley, A., Conway, M. and Aggleton, J. (eds) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.

Graham J. Hitch

---

## 1 Introduction

**Working memory** refers to our ability to co-ordinate mental operations with transiently stored information during cognitive activities such as planning a shopping trip or reading a newspaper. This chapter begins with a brief discussion that places the concept of working memory within the context of memory as a whole, then moves on to deal with distinctions between the concepts of working memory, Short-Term Memory (STM) and Long-Term Memory (LTM). Having done this, we are in a position to consider the architecture of working memory, that is, the unchanging features that account for its operation in different cognitive activities. We shall see that – in common with many other aspects of the cognitive system – identifying structure is no trivial task. The discussion is organized around the influential account of working memory presented by Baddeley (1986), tracing some of the developments in this model in the light of new evidence and noting alternative accounts where appropriate. The material covered has been chosen to illustrate the increasing diversity of phenomena that are seen as relating to working memory and includes evidence from laboratory experiments, individual differences, normal and abnormal development, neuropsychology and neuroimaging. We go on to focus in more detail on the particular topic of phonological working memory and vocabulary acquisition, where the convergence of different kinds of evidence is particularly striking. Finally, we take a short look at recent developments in computational modelling that attempt to make theories of working memory more precise. Overall, we shall see that, although we are beginning to understand more about working memory, many questions still have to be answered.

### 1.1 Human memory as a multifaceted system

When someone tells us they have a poor memory, they may be referring to any of a range of specific problems. For example, they may have difficulties in recalling past events, remembering to do things, or perhaps retrieving facts or names. In everyday life we tend to talk about memory as if it is a single faculty. However, there are many grounds for thinking that memory is multi-faceted, made up of a number of separate but inter-linked systems (see Chapter 8). Probably the oldest theoretical distinction of this kind is between a system for holding information over long periods of time and a system that deals with information over much shorter intervals, of the order of seconds or at most a few minutes. STM refers to our ability to retain temporary information over such intervals, as in looking up a telephone number and then dialling it. Working memory is a related concept but as our earlier examples of reading and planning make clear, it goes beyond the mere retention of information. More specifically, working memory keeps track of transient information and co-ordinates mental operations in a variety of cognitive tasks.

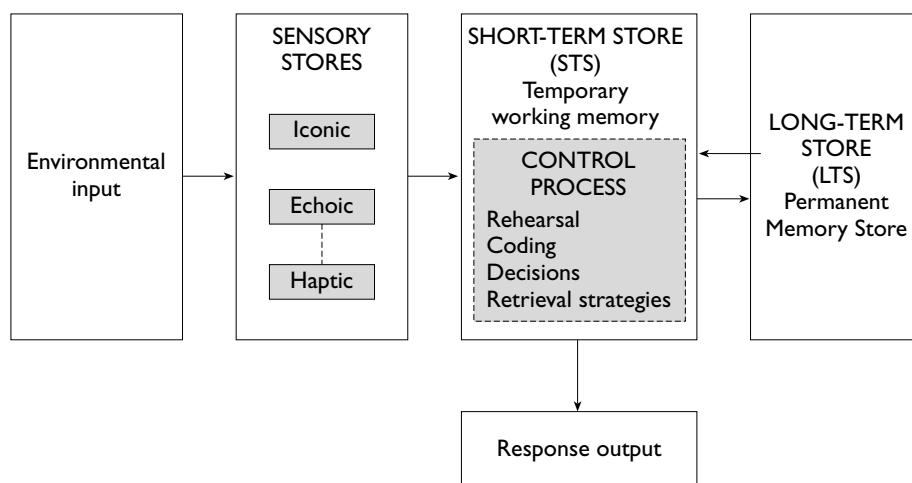
The classic illustration of working memory in action is complex mental arithmetic, where we typically break the task down into a series of operations. For example,  $26 + 37$  might be broken down into the stages  $20 + 30 = 50$  and  $6 + 7 = 13$  and  $50 + 13 = 63$  in order to get the answer. It can be seen here how the various stages have to be co-ordinated, and how early stages generate transient information that has to be maintained for eventual use in later stages. Experimental studies show that errors of mental arithmetic are mainly due to forgetting transient information during delays imposed by the sequencing of operations (Hitch, 1978). Written calculation overcomes this limitation of working memory by providing a durable external record. Other everyday examples of situations placing demands on working memory are talking to a group of unfamiliar people while trying to remember their names or taking notes while following a presentation. In such cases the combined demands of attending to mental operations while remembering transient information can cause difficulty and may result in errors, suggesting that working memory has a limited capacity. In order to discuss working memory in greater detail, it is necessary to sharpen the distinction between it and STM. This will be done in Section 1.3, but, in order to get closer to the roots of this distinction, we need first to go back to the origins of the historically earlier distinction between STM and LTM.

## 1.2 Distinction between short-term and long-term memory

Although William James first introduced the concept of ‘primary memory’ in 1890, it was not until the 1960s that an interest in memory over brief intervals of less than a minute became firmly established. Memory researchers at that time were pre-occupied with the question of whether or not human memory is a unitary mental faculty, as a number of different kinds of evidence were emerging that pointed to the idea of separate systems for short-term and long-term recall. One of these was evidence that memory for verbal stimuli has different properties over short and long intervals. For example, Baddeley (1966a) showed that immediate recall of a list of briefly presented words is poor when the items are phonemically similar to each other (e.g. share the same vowel, as in *man*, *can*, *cad*, etc.) but is unaffected when they are semantically similar (e.g. share the same meaning, as in *huge*, *big*, *large*, etc.). However, when the same materials are presented more than once and memory is tested after a longer retention interval, the accuracy of recall is lower for semantically similar items and is unaffected by phonemic similarity (Baddeley, 1966b). These observations pointed to two separate storage systems that code information in different ways. Information in STM is held in an acoustic or speech-based form whereas information in LTM is coded in terms of its meaning. Other evidence showed that the rate of forgetting briefly presented stimuli was unusually rapid when compared with forgetting rates for better-learned material, consistent with the idea that STM is much more labile than LTM (Brown, 1958). Over and above these observations, it had been known for quite some time that the so-called ‘span of immediate memory’ is limited to just a few items, whether these are digits, letters or words (e.g. Miller, 1956). Memory span is the longest sequence

that can be recalled accurately after a single presentation. The low limit on span suggested that STM can be distinguished from LTM on the grounds of its limited capacity.

So compelling was all this evidence at the time that several two-store models of memory were proposed. Reflecting this unanimity, their common features were referred to as the ‘modal’ model (Murdock, 1967). The main assumptions of this model were (1) that STM is a limited-capacity store of short duration, (2) that control processes, such as subvocal rehearsal, can be used to maintain information in STM, and (3) that information in STM is gradually transferred to LTM. Atkinson and Shiffrin (1971) provide the best known example of this type of account (see Figure 9.1).



**Figure 9.1** The Modal model of memory, redrawn from Atkinson and Shiffrin (1971). Note how information has to pass through the short-term store in order to access the long-term store. Note also that the sensory stores are not discussed in the text. They are very short-lived and are specific to the various sensory pathways that feed information into the short-term store

Source: Atkinson and Shiffrin, 1971

You will see from the diagram that Atkinson and Shiffrin (1971) labelled their short-term store as a working memory that serves other functions besides acting as a temporary store. These functions include the regulation of control processes such as rehearsal or retrieving information from LTM. Note that control processes are optional and are conceptually different from involuntary, automatic processes. At about the same time that the Atkinson and Shiffrin model was enjoying its popularity, numerous other authors argued that the transient storage provided by STM was crucial for cognitive activities such as sentence comprehension or problem-solving. In other words, there was a general assumption that STM behaves as some form of working memory. You can gain some insight into the plausibility of supposing that these activities require keeping track of temporary information within a stream of ongoing mental operations by trying one for yourself (see Box 9.1 on the comprehension of garden-path sentences).

**9.1****Methods****Understanding ‘garden-path’ sentences**

Garden-path sentences are sentences that lead the comprehender ‘up the garden path’ towards an incorrect interpretation, as in *We painted the wall with cracks* (see Chapter 6). It is the ambiguity of such sentences that makes them difficult. One explanation assumes that multiple interpretations of ambiguous sentences are held in working memory (Just and Carpenter, 1992). Just and Carpenter support their view with evidence that individuals with low working memory capacity are less able to maintain multiple interpretations than individuals with high working memory capacity. However, an alternative theory is that comprehension draws on more specialized resources than working memory (e.g. Caplan and Waters, 1999).

Despite the fact that the modal model captures some important insights, the consensus it reflected was somewhat fleeting. One concern was whether the various strands of evidence for distinguishing STM and LTM converged on a coherent account. For example, different ways of estimating the capacity of the short-term store gave quite different answers and the reasons for this were unclear. The immediate consequence of this challenge was a resurgence of interest in LTM (e.g. the ‘levels of processing approach’ proposed by Craik and Lockhart, 1972) rather than attempts to revise and refine the concept of the short-term store. Another concern was whether the short-term store does in fact act as a working memory. One example causing difficulty for this position was some intriguing neuropsychological evidence from a patient known in the literature as KF who sustained brain damage as a result of a road accident (Shallice and Warrington, 1970). KF’s auditory digit span was only two items which is way below the normal range of seven plus or minus two items identified by Miller (1956). However, despite having such a severe deficit, KF performed normally on tests of long-term learning and memory, he had normal intelligence and no major difficulties in understanding spoken language (Shallice and Warrington, 1970). In one respect KF’s pattern of memory performance was consistent with the modal model: it could be explained in terms of selective damage to his STM while his LTM was intact. Moreover, the fact that damage to part of the brain could have this effect suggested a separate neuroanatomical localization of the short-term store. However, the absence of a general impairment in KF’s learning, comprehension and reasoning presents obvious difficulties for the idea that STM acts as a working memory that is necessary for supporting such activities.

**1.3 Working memory as more than STM**

Given difficulties such as those presented by KF, Baddeley and Hitch (1974) made an empirical investigation of whether STM does indeed act as a working memory. One technique they used was the **dual-task paradigm** in which people perform two tasks at the same time. The logic of this paradigm is that two tasks will interfere with one another if they require access to a common resource and if their combined demands exceed its capacity. Baddeley and Hitch (1974) examined the effect of requiring people to perform an irrelevant STM task at the same time as a cognitive task that involved either reasoning, comprehending language or learning new

information. For example, in one experiment people carried out a verbal reasoning task while remembering sequences of random digits (see Box 9.2 for an outline of the experimental procedure). Reasoning was impaired when the STM load was increased by making the digit sequences longer. Similar results were obtained when the cognitive task was either comprehending prose or learning a list of words for free recall. Baddeley and Hitch (1974) drew two main conclusions from these observations. First, the finding that an irrelevant STM task interferes with a range of cognitive tasks is consistent with the idea of a common working memory system that combines temporary information storage with ongoing mental operations. Second, working memory goes beyond the concept of STM. Thus, even when the load on STM approached memory span, and therefore ‘filled’ short-term storage capacity, there was no catastrophic breakdown in concurrent cognition. This suggests the idea that working memory includes an additional resource that is not shared with STM.

## 9.2

### Research study

#### **Studying the effect of an irrelevant memory load on verbal reasoning**

The verbal reasoning task used by Baddeley and Hitch (1974) involved deciding whether a sentence gave a true or a false description of the order of a letter pair. Examples are, *A precedes B – AB* (true), and *B does not follow A – AB* (false). Varying the verb, the grammar, the letter order and the truth-value of the answer gave a total of 32 problems of varying difficulty. Each problem was shown individually, performance being measured by the speed and accuracy of pressing ‘true’ and ‘false’ response keys.

One experiment involved a comparison between the effect on reasoning of concurrently repeating a sequence of six random digits and counting repeatedly from one to six. The rationale was that a sequence of six random digits is close to the span of immediate memory, whereas the counting sequence is stored in long-term memory. Repeating random digits slowed solution times in the reasoning task, relative to a control condition, but the counting task had very little effect. Furthermore, the interference produced by random digits was greater for the more difficult versions of the reasoning task. The conclusion Baddeley and Hitch (1974) drew was that reasoning and short-term retention compete for a limited-capacity ‘workspace’ that can be flexibly allocated to either the storage demands of the memory load or the processing demands of the reasoning task.

Further evidence for a distinction between STM and working memory came from studies of individual differences. The logic behind this approach is that if two tasks involve similar underlying psychological processes, a person who performs well on one should do well on the other. In statistical terms, the two abilities should be positively correlated. In an influential study, Daneman and Carpenter (1980) argued that standard measures of STM, such as word span and digit span, tax storage capacity but do not assess the capacity to combine storage with ongoing processing operations. In order to provide a better assessment of the latter, and therefore of

working memory, Daneman and Carpenter devised a novel reading span task. In this task, participants were required to read aloud a set of unrelated sentences and immediately afterwards to recall the last word of each sentence. Box 9.3 gives further information about the procedure. As you will see if you try it for yourself, the task rapidly becomes very demanding as the number of sentences increases. To assess the limit on reading span, Daneman and Carpenter (1980) prepared three sets each of two, three, four, five and six sentences. Participants were presented with increasingly longer sets of sentences until they failed all three sets at a particular level. An individual's reading span was taken as the maximum level at which they were correct on at least two of the three sets. The procedure is analogous to standard measures of STM span in that it assesses the longest sequence of items that can be maintained over a short interval. However, in reading span, the items have to be remembered at the same time as performing the processing operations required for reading sentences, whereas in STM span there is no simultaneous processing requirement.

### 9.3

### Research study

#### Procedure for determining reading span

The materials for Daneman and Carpenter's (1980) reading span task were a set of unrelated sentences, each of which was typed on a separate card. The two examples they gave are:

*When at last his eyes opened there was no gleam of triumph, no shade of anger.*

*The taxi turned up Michigan Avenue where they had a clear view of the lake.*

Cards were arranged in sets of two, three, four, five and six sentences, there being three instances of each set-size. Participants were shown one card at a time and read it aloud at their own pace, starting at set-size two. The second card was presented as soon as the first was read. A blank card signalled recall of the final word on each card in their order of occurrence (i.e. *anger, lake* in the above example of set-size two). Three trials were given at each set-size, and set-size was increased until all three trials at a particular level were failed. At this point testing was ended. Reading span was taken as the level at which the participant was correct on two out of three sets. As with memory span, there are many variants on this basic procedure.

Daneman and Carpenter (1980) compared reading span with word span as predictors of reading comprehension skills in a group of college students. Reading comprehension was measured in three ways: fact questions, pronoun questions and verbal SATs (see Table 9.1). It turned out that reading span was a very good predictor of all three measures and a much better predictor than word span. Daneman and Carpenter went on to show that a listening span measure gave similar results, showing that the correlation is not specific to reading. They interpreted their findings as showing that working memory capacity is an important source of individual differences in language comprehension, the key characteristic of working memory being combining temporary storage with information processing, in line with the approach taken by Baddeley and Hitch (1974).

**Table 9.1** Correlations between spans and various measures of reading comprehension

	Reading comprehension measure		
	Fact questions	Pronoun questions	Verbal SAT
Reading span	.72	.90	.59
Word span	.37	.33	.35

Source: Daneman and Carpenter, 1980, experiment 1

You are probably already well aware that correlations can be interpreted in many ways. Thus, a criticism often made of Daneman and Carpenter is that their correlations might be an artefact of similarities in processing operations in the various tasks they used. Reading span, listening span and language comprehension all involve language processing whereas word span does not. The potential force of this criticism is substantial and called into question whether Daneman and Carpenter's results have anything to do with working memory as a general-purpose resource. To address it, other investigators have looked at patterns of correlation using different measures of working memory span to which the criticism does not apply. For example, Turner and Engle (1989) devised an operation span task in which participants solved sets of arithmetical calculations. After each calculation was completed a word was presented and at the end of the set all the words had to be recalled. Operation span was the limit on how many words could be recalled under these conditions. Turner and Engle (1989) found that operation span was a superior predictor of reading comprehension than was standard STM span, despite involving dissimilar processing operations. Their results therefore provide support for the idea of a general working memory system that is common to a range of different activities involving the combination of information processing with temporary storage. Subsequent work by Engle *et al.* (1999b) has expanded this picture by showing that working memory span is more closely related to general intelligence than is STM.

## Summary of Section 1

---

- Human memory can be seen as a multifaceted system whose distinct components have different characteristics and functions.
- An important distinction is that between a transient, limited-capacity, STM system and a more stable LTM system.
- Atkinson and Shiffrin (1971) suggested that STM acts as a working memory responsible for a variety of control processes.
- Baddeley and Hitch (1974) explored and expanded this idea and concluded that STM is better regarded as a component of working memory.
- Converging evidence that working memory and STM are not identical comes from studies of individual differences, e.g. Daneman and Carpenter (1980) found that reading span was much better than word span for predicting verbal abilities.

## 2 The structure of working memory

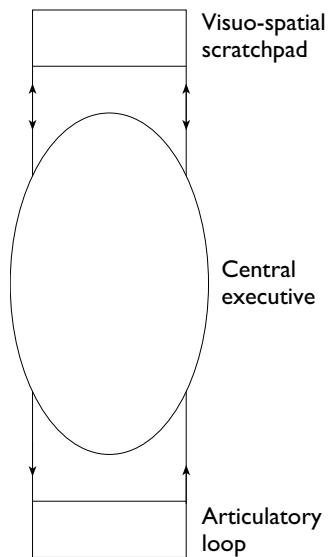
We have seen some of the evidence suggesting that working memory differs from STM, but so far little about how it differs beyond referring to evidence that working memory includes STM. This section covers the structure of working memory in more detail.

### 2.1 A multi-component model

In their original investigation, Baddeley and Hitch (1974) studied whether irrelevant STM loads affected reasoning, language comprehension and list learning. Their aim was to examine whether these cognitive activities involve the same limited capacity as STM. Although high STM loads did cause interference, people could retain low loads of two or three items without much disruption to the primary task. This observation was seen as consistent with the suggestion that working memory can be partitioned into two components, one that can hold small amounts of temporary information and another that is more concerned with cognitive processing. In further experiments Baddeley and Hitch (1974) looked at the effects of varying the phonemic similarity of the materials in reasoning and comprehension tasks. Adverse sensitivity to phonemic similarity is a characteristic feature of STM (see Section 1.2), and showing that reasoning and comprehension are also sensitive would suggest that they share a common factor. In the reasoning task, subjects were asked to verify relationships such as 'A is not preceded by B - AB', where the letters used were either phonemically similar (e.g. TD) or dissimilar (e.g. MC). In the comprehension task, subjects were asked to say whether the words of a sentence were presented in a meaningful or jumbled order. The words either rhymed (e.g. *Red headed Ned said Ted fed in bed*) or did not rhyme (e.g. *Dark skinned Ian thought Harry ate in bed*). The results showed that phonemic similarity did disrupt reasoning and comprehension, but only somewhat mildly.

To account for their results, Baddeley and Hitch (1974) assumed that one of the components of working memory is a limited-capacity, speech-based store capable of storing two to three items. This subsystem was described as an **articulatory rehearsal loop** and can be viewed as roughly equivalent to the earlier concept of STM (more detail about the articulatory loop is given in Section 2.2). The articulatory loop could be used to store small memory loads during cognitive tasks and was responsible for the effect of phonemic similarity on performance. The second component was described as a **central executive**, responsible for the control and co-ordination of mental operations in a range of activities including but extending beyond reasoning, comprehension, learning and memory. The executive was seen as a limited-capacity workspace that can be flexibly allocated to control processes or temporary information storage, depending on the nature of the task in hand. Thus a small irrelevant memory load could be stored in the articulatory loop without taxing the central executive, but a larger memory load would take up extra resources in the executive. Given a limit on the capacity of the workspace, this theoretical account maintains that there will be a trade-off such that fewer resources are available to support processing operations when temporary storage demands increase.

In reflecting on their results, Baddeley and Hitch noted that the tasks they had investigated were all primarily verbal. The question arose as to whether tasks involving visual memory and visual imagery also draw on working memory and, if so, how. The information available from dual-task studies indicated that combining two visuo-spatial activities (such as tracking a moving object while performing a mental imagery task) or combining two verbal activities is more difficult than combining one of each. This observation suggests there are separate resources specialized for dealing with verbal and visuo-spatial information. Nevertheless, as there is some mutual interference when a visuo-spatial and a verbal task are combined, the data are also consistent with the involvement of a common resource. One way of accounting for these observations is to assume that the central executive controls visual and verbal tasks and that there is a separate subsystem for storing visuo-spatial information, analogous to the articulatory loop. This tripartite model, in which the extra subsystem is referred to as the visuo-spatial sketchpad, was developed further by Baddeley (1983; 1986) and is illustrated in Figure 9.2.



**Figure 9.2** The structure of working memory

Source: based on Baddeley, 1983

the dual-task methodology, Baddeley and Lieberman (1980) made the interesting observation that use of a visual imagery mnemonic was disrupted by a spatial task (tracking a moving loudspeaker while blindfold) but not by a visual task (detecting changes in the brightness of a blank field). This pattern was not observed when the mnemonic strategy was rote rehearsal instead of imagery, suggesting it was not a function of the relative difficulty of the spatial and visual interfering tasks. Baddeley and Lieberman (1980) interpreted their results as evidence that mental imagery is spatial rather than visual.

Unfortunately there is not the space to deal with visuo-spatial sketchpad in the detail it deserves. However, one interesting observation is that neurological patients can show selective impairments in visuo-spatial STM and imagery tasks suggestive of a separate brain location for visuo-spatial function. Corsi span is a test of visuo-spatial STM in which a set of nine identical cubes is mounted at haphazard locations on a horizontal board. The experimenter points to a selection of cubes and the task is to reproduce the sequence immediately by pointing. Sequence length is progressively increased and the limit beyond which performance breaks down defines span. De Renzi and Nichelli (1975) found that Corsi span and auditory digit span could be impaired independently in patients with different lesions. Evidence such as this is strongly indicative of a separate, non-verbal store. Such a store may underpin the use of visual coding to remember verbal items. The formation of mental images as mnemonics to aid recollection has a long history going back at least as far as Ancient Greece. Using

However, this somewhat counterintuitive conclusion does not generalize to all forms of imagery. Hitch, Brandimonte and Walker (1995) studied people's ability to perform an imagery task in which they were shown two separate line drawings. They then had to superimpose mental images of the drawings in order to reveal a novel percept. For example one drawing looked like two ice cream cones and the other showed a curved line whose ends coincided with the locations of the tops of the cones. When mentally superimposed, the drawings combined to reveal a skipping rope. Hitch *et al.* (1995) found that imagery performance was better when the drawings were visually congruent (i.e. both consisted of a black figure on a white ground) than when they were incongruent (i.e. their contrasts were reversed). Thus in this particular imagery task, there is clear evidence that the images preserve information about visual appearances. It is interesting to note in passing that if you were able to 'see' the skipping rope in your mind's eye after reading the above descriptions, you achieved this using conceptually-driven images rather than the perceptually-driven images studies in Hitch *et al.*'s (1995) experiment. The visual characteristics of the two types of image are not necessarily the same.

In a review of visuo-spatial working memory, Logie (1995) suggested that there are separate spatial and visual systems, such that a spatial movement system can be used to rehearse the contents of a visual store. This proposal corresponds to a visuo-spatial analogue of the articulatory loop. However, the full story about imagery and working memory is still unfolding and may be considerably more complex. For example, Smyth and Waller (1998) asked rock climbers to imagine tackling familiar routes while performing a variety of secondary tasks designed to disrupt their ability to use visual, spatial or kinaesthetic information. The results implicated multiple forms of representation and pointed to the complexity of imagery for skilled movement.

In conclusion, the work of Baddeley and Hitch (1974) led to a tripartite model of working memory that was subsequently developed by Baddeley (1986). This model appears to have been the first substantive account of working memory and has been influential within the field. However, an increasing number of alternative accounts has emerged subsequently, many of which are described in a recent volume edited by Miyake and Shah (1999). Several theoretical issues divide these approaches. One of the principal questions concerns the relationship between working memory and LTM. Baddeley and Hitch (1974) assumed that the two were separate systems. However, a number of authors take a different view, maintaining that working memory corresponds to an activated region of LTM (e.g. Ericsson and Kintsch, 1995; Cowan, 1988). Part of the motivation for this alternative approach comes from the effects of a person's degree of knowledge in a specific domain on their working memory capacity in that domain. For example, chess experts display superior working memory skills when given tasks within the chess domain. There is much more to be discovered about effects such as these and their interpretation. However, it is interesting to note that Cowan (1988) still assumes a separate executive system, making the difference of view one concerning the nature of back-up storage (that is, specialized buffer stores versus activated LTM) (see also Engle *et al.*, 1999a). The idea of specialized buffer stores has also been challenged by the

work of Jones (see Section 2.3.4). In the remainder of this chapter we stay within the Baddeley and Hitch framework for the purpose of organizing the discussion, raising problems for it where appropriate. We begin with the relatively well-specified concept of the articulatory loop, before moving on to the central executive, the most important but still least well understood aspect of working memory.

## 2.2 Phonological working memory

One reason the articulatory loop is relatively well understood is the existence of a cluster of experimental manipulations that affect its operation. We have already encountered one of these, namely the phonemic similarity of items presented in tests of immediate recall (see Section 1.2). A second variable was the word length of the items. In an important series of experiments, Baddeley, Thomson and Buchanan (1975) showed that the limit on STM span for verbal stimuli was not a fixed number of items or chunks, as Miller (1956) had claimed. They showed instead that memory span varies with the length of the items, being higher for shorter items (e.g. *harm*, *wit*) than for longer items (e.g. *university*, *hippopotamus*). Box 9.4 describes one of their procedures and results. One of many interesting observations was that there was a systematic relationship between how many words could be recalled and the time it took to say them out loud. Thus, people could recall the number of words that could be spoken in about two seconds. This is consistent with the idea of a rehearsal loop in which rehearsing items refreshes their decaying memory traces. Longer words take longer to rehearse so fewer can be refreshed within two seconds, the time limit set by the rapidity of the decay process. Baddeley *et al.* (1975) also examined individual differences and found that faster speakers tended to recall more information than slower speakers. This is consistent with the model if one assumes that faster speakers can rehearse more rapidly. The model could also account for the phonemic similarity effect, as a given amount of decay would have a greater effect on the ability to discriminate the memory traces of items that share phonological features. To appreciate this point, suppose you have been presented with the sequence of phonemically similar letters BTG to recall. If, as a result of partial forgetting of the third item, you could only remember that it contained an /e/ sound, this would not be very helpful as it leaves many options open. Compare this with a sequence of dissimilar items such as RJQL, where being able to remember that the third item had a /u/ sound would be of much more help.

This model of the articulatory loop was also able to explain the results of dual-task experiments in which immediate serial recall was combined with **articulatory suppression** (a secondary task involving the repetition of a redundant and irrelevant word such as *the the the*). Articulatory suppression simply requires the participant to repeat a word over and over again. This low-level **secondary task** is intended to occupy the articulatory loop with irrelevant (but unavoidable) activity, so that performance on the **primary task** has to manage without the assistance of the articulatory loop (or at least without a large part of its functioning). Baddeley *et al.* (1975) found that articulatory suppression disrupted recall, consistent with it disrupting use of the articulatory loop. Suppression also removed differences between the recall of longer and shorter words and between phonemically similar and dissimilar items. These further effects are also consistent with disruption of the loop. However, the effects of word length and phonemic similarity only disappeared

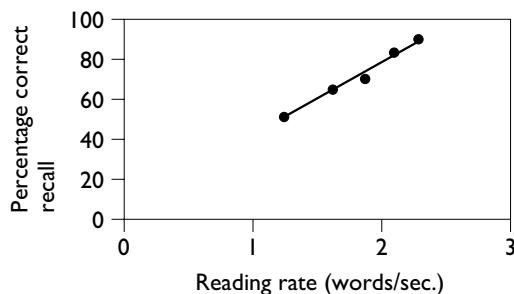
## 9.4

## Research study

### The word-length effect

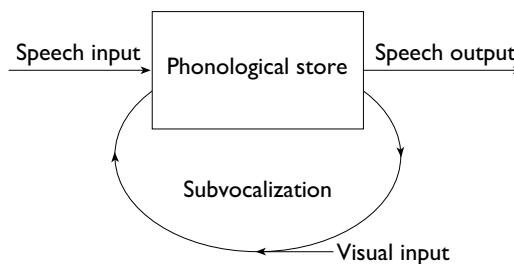
In one of their experiments, Baddeley *et al.*, (1975) constructed five pools of 10 words of one, two, three, four or five syllables. The pools were matched for semantic category and familiarity. To illustrate, the one-syllable pool included *Stoat*, *Mumps*, *School*, *Greece*, and corresponding items in the five-syllable pool were *Hippopotamus*, *Tuberculosis*, *University*, *Yugoslavia*. Ten lists of five words were made up of random permutations within each pool. The lists were presented in a random order, words being shown one after another at a two-second rate. Immediately after list presentation, participants spoke their recall. In a second part of the experiment, reading rate was measured. This was achieved by timing participants reading aloud a typed list of the words in each pool as quickly as they could.

The results showed that the percentage of words recalled dropped as the number of syllables increased. Moreover, as the graph shows (see Figure 9.3), the plot of percentage correct recall against articulation rate formed a straight line. The slope of the line was about two seconds, demonstrating that, the faster a person can say a list of words out loud (that is, the faster they can rehearse), the more effective they prove in subsequently recalling those words.



**Figure 9.3** Results obtained by Baddeley *et al.* (1975). Percentage of words recalled is plotted as a function of the rate at which the same words could be read aloud, for five different word-lengths. The point furthest to the right corresponds to one-syllable words, the next point to the left represents two-syllable words, and so on

when items were presented visually and not when they were presented auditorily. This unexpected effect of presentation modality was for some time something of a puzzle. The position was eventually clarified in experiments carried out by Baddeley *et al.*, (1984) where suppression was continued during recall as well as item presentation. Under these conditions, suppression removed the word-length effect for auditory items, but still did not remove the phonemic similarity effect. Baddeley *et al.* (1984) explained these results in terms of a modified theoretical account in which the articulatory loop is seen as consisting of a decaying phonological store (the locus of the phonemic similarity effect) and a control process of subvocal rehearsal (the locus of the word-length effect) (see Figure 9.4). According to this account, spoken stimuli access the loop automatically whereas visual inputs have to be verbally recoded, an optional control process that involves **subvocalization**. Suppression eliminates the word-length effect for both visually and auditorily



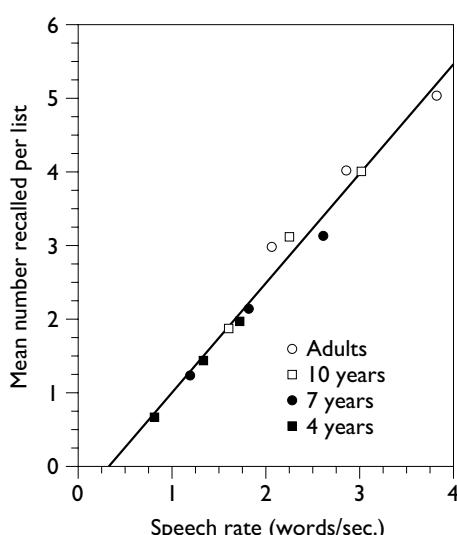
**Figure 9.4** The structure of the phonological loop, according to the ideas developed by Baddeley *et al.*, 1984

presented stimuli by disrupting rehearsal, but only eliminates the effect of phonemic similarity for visually presented stimuli as only this type of stimulus requires verbal recoding. In this way, specification of different pathways by which visual and spoken stimuli access the loop explains an otherwise obscure pattern of findings. Nowadays, it is more common to use the term **phonological loop** to refer to this more developed, two-component account of the articulatory loop. The next section shows how this model of the phonological loop generates useful insights into developmental changes in verbal STM as children grow up.

### 2.2.1 Developmental and cross-linguistic differences

The two-part model of the phonological loop is interesting in a number of different ways. Not least is that the model can be applied to phenomena outside its initial scope. One example is the developmental growth of memory span during childhood, for which many competing explanations have been proposed (Dempster, 1981). Thinking in terms of the phonological loop model suggests it would be informative

to measure children's recall of lists of words of different lengths and the speed at which they can articulate the words, as in Baddeley *et al.*'s (1975) study of adults. The results of doing this are quite striking. As children's ages increase, their average level of recall increases in proportion to the rise in their average speech rate (Nicolson, 1981; Hulme *et al.*, 1984; see Figure 9.5). Furthermore, the size of the word-length effect in children of different ages reflects the time it takes to articulate words of different lengths. Finding such a clear empirical relationship is informative and suggests a possible explanation for the developmental growth in memory span. Thus, if older children can rehearse faster, then they can maintain more items within the approximately



**Figure 9.5** Results obtained by Hulme *et al.*, 1984. Percentage of words recalled is plotted as a function of their speech rate for three different word-lengths and four age-groups

two-second time-limit set by trace decay in the phonological store. Notice, however, that this is a causal interpretation of a correlation and therefore difficult to prove conclusively. Yet another phenomenon to which the concept of the phonological loop has been applied is cross-linguistic differences in digit span. For example systematic differences in mean digit span in English (7.2 digits), Spanish (6.4 digits), Hebrew (6.5 digits) and Arabic (5.8 digits) that would otherwise be difficult to explain turn out to be highly correlated with differences in the rates at which the digits can be articulated in these languages (Naveh-Benjamin and Ayres, 1986).

The phonological loop model has prompted further discoveries about developmental change. One of these discoveries concerns the effect of word length when children remember a sequence of stimuli presented as either spoken words or nameable pictures. Older children aged around seven upwards show the standard tendency for poorer recall of items with longer names for both types of stimulus, but younger children show this only for spoken stimuli (Hitch *et al.*, 1989). Moreover, when recalling nameable pictures, younger children find it harder when the pictures are visually similar to one another whereas older children find it harder when the names of the pictures are phonemically similar (Hitch *et al.*, 1988). These observations are consistent with the assumption that auditory stimuli gain automatic access to the loop but that phonological recoding is necessary for visual stimuli. They suggest further that the process of recoding is somewhat slow to develop and that younger children are more reliant on visuo-spatial working memory. Subsequent research has confirmed the developmental progression from visual to phonological coding and suggests that it is related to learning to read, being markedly delayed in dyslexic children (Palmer, 2000a; 2000b).

### 2.2.2 The irrelevant speech effect

Yet another application of the phonological loop model was to explain why the presence of background speech disrupts STM for visually presented verbal stimuli. Salamé and Baddeley (1982) showed that having to ignore irrelevant speech was more interfering than ignoring irrelevant noise, leading them to suggest that unattended speech enters the phonological store whereas non-speech sounds do not. Consistent with such an interpretation, blocking people's ability to verbally recode visual stimuli by having them suppress articulation removes the disruptive effect of irrelevant speech (Salamé and Baddeley, 1982). However, this account has been challenged by evidence that unattended non-speech sounds can cause interference, and that the amount of interference is determined by the same factors as for speech. One such common factor is that steady-state streams (where the irrelevant stimuli remain the same) cause less disruption than changing-state streams (where the irrelevant stimuli vary over time) (Macken and Jones, 1995). Such observations have been used to question the assumption that irrelevant speech has an effect that is specific to the phonological loop. They suggest a broader explanation of the interference caused by irrelevant sounds in terms of general memory mechanisms that are not specific to the verbal domain. The irrelevant speech (or sound) phenomenon has developed into an area of considerable controversy (see for example, Baddeley and Larsen, 2003). However, the two theoretical approaches are not mutually exclusive and it may be, for example, that irrelevant speech affects both a general mechanism (e.g. for serial ordering) as well as the phonological store.

### 2.2.3 Neural basis

Before closing this part of the discussion, we shall consider some evidence about the neural basis of the phonological loop. An obvious challenge for any model is to explain neuropsychological cases of selective impairment of memory span of the type demonstrated by Shallice and Warrington (1970). Vallar and Baddeley (1984) made a detailed investigation of one such patient, known as PV. Following a stroke that left her with damage that included her left parietal cortex, PV's auditory digit span was reduced to only two items. However, her other abilities were relatively unimpaired. For example, her speech was fluent and her rate of articulation was normal. Vallar and Baddeley (1984) found that PV's memory span for spoken sequences was poorer when the items were phonemically similar but was unaffected by their word length. They interpreted these observations as indicating that her phonological store was damaged. Thus, if the store was functioning at a reduced level, spoken inputs would access it automatically and immediate recall would be sensitive to the phonemic similarity of the items. However, given a damaged phonological store, PV would not find subvocal rehearsal a useful strategy. Hence, she would not show the normal word-length effect. Like other patients of this type, PV's memory span for visually presented verbal stimuli was higher than her auditory span. Moreover, her visual span was unaffected by either phonemic similarity or word length of the materials. These observations suggest that PV may have been relying on visuo-spatial working memory to remember visual stimuli. There may be an interesting parallel to be drawn here with children's reliance on visuo-spatial working memory for remembering visual stimuli early on in development when, albeit for different reasons, their ability to use the phonological loop is restricted (see Section 2.2.1).

Research on patients raises the question of the neuroanatomical localization of the phonological loop. Neuroimaging techniques provide the opportunity to study this in the normal brain. In an early study, Paulesu *et al.*, (1993) investigated which areas of the brain are active in tasks thought to involve the phonological loop. Such experiments depend on a subtraction logic whereby brain activation observed in one experimental task is compared with that in another. By arranging that the two tasks differ solely in the process of interest, the neural activation specific to that particular process can be obtained by subtraction. This of course is not as simple as it sounds and typically involves making theoretical assumptions about the tasks under consideration. Paulesu *et al.* (1993) compared activation patterns in a verbal memory task requiring storage and rehearsal, a rhyme judgement task that required rehearsal but not storage, and a control task requiring neither storage nor rehearsal (Box 9.5 (overleaf) describes the experiment in more detail). The results suggested separate localisation of storage and rehearsal, consistent with the theoretical distinction between these two aspects of the phonological loop. Furthermore, localization of the store to an area in the left parietal cortex corresponded approximately to the locus of damage in patients like PV. Other neuroimaging studies converge with – but also complicate – this simple picture, especially with regard to the involvement of other brain areas (e.g. Henson, 2001). My purpose here is merely to illustrate an early success in using the phonological loop model to guide the collection and interpretation of neuroimaging data.

**9.5****Research study****Neural correlates of the phonological loop**

Paulesu *et al.* (1993) used positron emission tomography (PET) to measure blood flow in different regions of the brain. This technique involves making an intravenous injection of radioactive water and then scanning the brain to record the spatial distribution of radioactivity. Scanning is performed during matched tasks that differ with regard to a feature of interest. Subsequent comparison of the two activation patterns allows brain regions associated with the feature of interest to be identified. (A similar logic applies to functional magnetic resonance imaging (fMRI) a more recent technique that does not involve radioactivity.)

Paulesu *et al.* compared brain activation patterns in phonological and non-phonological memory tasks. The phonological task involved showing a sequence of six consonants followed by a probe letter. Participants indicated whether the probe item had appeared in the sequence. The non-phonological memory task was identical except that the items were unfamiliar Korean characters. The two tasks were therefore closely matched, but only remembering consonants engaged the phonological loop. Subtracting activation patterns revealed that the consonant memory task was associated with increased blood flow in left hemisphere regions corresponding to Broca's area and the supramarginal gyrus of the parietal cortex (see Colour Plate 5).

A second comparison was between a rhyme judgement task and a shape judgement task. In the rhyme task participants saw a series of consonants and indicated whether each one rhymed with the letter B, which was always present. The shape task was identical except that the stimuli were Korean characters and the judgement was one of shape similarity. Previous research suggested that the rhyme judgement task would engage the subvocal rehearsal system but not the phonological store. Subtraction of the scans indicated that the rhyme task activated Broca's area, but not the left supramarginal gyrus. Thus, the subvocal rehearsal system can be identified with Broca's area and, by revisiting the subtraction for the memory tasks, the phonological store can be identified with the left supramarginal gyrus.

**2.2.4 Theoretical issues**

We have seen how a simple model of the phonological loop has proved productive in ways that extend beyond its initial remit. These applications cover a surprisingly extensive range that includes developmental and cross-linguistic differences, effects of irrelevant speech, cases of neuropsychological impairment and results of neuroimaging studies. The model has turned out to be remarkably successful – it has evidently ‘travelled well’. However, some of its limitations are steadily becoming more apparent, as in its explanation of the effects of irrelevant speech. Other recent evidence suggests that the word-length effect may not be due to differences in items’ spoken duration. Thus, there is little or no effect of word duration when the phonological complexity of items is carefully controlled (Lovatt *et al.*, 2000). In addition, developmental studies suggest that rehearsal is not necessary for the word-length effect. Specifically, children as young as four show a word-length effect when

recalling spoken stimuli, an age when it is generally agreed they have not acquired the ability to use rehearsal strategies (Hulme *et al.*, 1984). Other authors have shown that output delays are sufficient to cause word-length effects, without appealing to rehearsal (Brown and Hulme, 1995; Cowan *et al.*, 1992).

Whether the limitations of the phonological loop as a model count as falsifications is an interesting scientific issue that might send us back to the drawing board for an entirely new account. Some authors have taken this approach (Nairne, 2002). The alternative strategy is to revise the model to overcome its limitations, while at the same time preserving its original insights. We saw an earlier example of this in the elaboration of the account of the phonological loop to explain why the effects of articulatory suppression differ when the memory items are seen rather than heard (see Section 2.2). A more recent example is the effort to develop the concept of the phonological loop through more detailed computational modelling (see Section 4). It is probably too soon to say which of these strategies will be the more productive – a totally new approach or development based on the present model. Only time will tell. For the present we note that, despite its limitations, the phonological loop continues to provide a simple, usable framework for linking a robust set of psychological phenomena, and is still widely used. However, before continuing with further discussion of the phonological loop, we turn to the main aspect of working memory in the tripartite model: the central executive.

## 2.3 Executive processes

The central executive is, in general terms, responsible for controlling and co-ordinating mental operations in working memory. Baddeley and Hitch (1974) suggested that the functions of the executive included supervising slave stores such as the phonological loop and the visuo-spatial sketchpad, as well as interactions with LTM. However, as we shall see, more precise identification of executive functions is a matter of continuing debate. The executive is at once the most important component of working memory, the most controversial and the least understood. At various times it has been described as a ‘ragbag’ or an area of ‘residual ignorance’ and, in a recent review, Andrade (2001) referred to it as ‘problematic’. There are good reasons for these remarks. One is that the executive could be seen as merely a reinvention of the somewhat derided concept of the homunculus, a person inside the head. The well-known problem here is that of explaining what controls the homunculus without appealing to an infinite regress of homunculi. Another difficulty is that, at an intuitive level, executive processes clearly have links to our sense of conscious awareness. This is another difficult concept, with a long history of intractability (see Chapter 15 on consciousness). However, rather than allowing themselves to be put off by these problems, researchers have attempted to understand what aspects of executive control they can, with the long-term goal of steadily reducing the area of residual ignorance.

### 2.3.1 Central workspace

We read earlier how Baddeley and Hitch (1974) conceptualized the executive as a limited-capacity central workspace with resources that could be flexibly allocated to various combinations of mental operations and temporary information storage. We also saw how Daneman and Carpenter’s (1980) reading span task was designed as a

method for assessing the capacity of such a workspace. Thus, given the assumption that resources for processing and storage trade off against each other, reading span can be interpreted as a measure of residual storage capacity when the workspace is also occupied in supporting reading processes. However, further evidence is needed to confirm that it is useful to think of the limited span of working memory as reflecting the capacity of a workspace or ‘mental blackboard’.

Several investigators have tried to examine more precisely what limits the span of working memory in tasks such as reading span and listening span. Given that the number of items in store increases from the start to the end of a trial, the workspace hypothesis predicts a corresponding decline in the resources available to support processing. This would follow from the trade-off between resources within the workspace. Towse *et al.*, (1998) tested this prediction by studying the performance of children on reading span, operation span and counting span in a series of parallel experiments. (Counting span involves presenting a set of visual displays showing random dots that must be counted. At the end of the set, the totals must be recalled and counting span is the maximum number of totals successfully recalled.) The results gave no clear support for the prediction, in that there was no systematic change in the speed of processing operations within trials. Towse *et al.* (1998) also entertained an alternative hypothesis according to which, rather than sharing attention between processing and storage, children switch attention back and forth between processing and storage. Thus, in reading span for example, children might read a sentence, store the final word, read the next sentence, store its final word and so on. According to this ‘task-switching’ account, reading span is limited by the rate of forgetting sentence-final words during the time intervals spent in reading. This is similar to the way in which errors in mental arithmetic were explained (Hitch, 1978) and is quite different from the resource-sharing account. To test the task-switching hypothesis, Towse *et al.* (1998) manipulated the time intervals over which information had to be stored in different conditions in which the total amount of processing was held constant. This was achieved by altering the order of presentation of the items within a set, some of the items being designed to take longer to process than others. In line with the prediction from task-switching, spans were lower when the intervals over which information had to be maintained were longer. This was true for all three tasks, reading span, operation span and counting span, suggesting a result of some generality. Subsequent research confirmed this by showing that manipulating the order of presentation of items has similar effects in adults (Towse *et al.*, 2000).

Other investigators have also found an effect of the length of the intervals devoted to processing operations in working memory span tasks, but have shown also that span is lower when the operations themselves are more complex (Barrouillet and Camos, 2001). Moreover, Hitch *et al.*, (2001) found some evidence for a trade-off in the form of a weak tendency for processing operations to become slower as storage load increased. Effects such as these lead us towards a mixed model that involves both attention switching and resource sharing. Further evidence suggests that other factors may also be involved in limiting working memory span. For example, de Beni *et al.* (1998) found that individuals with low spans made more intrusion errors where they erroneously recalled items from previous trials. This observation suggests that the ability to inhibit potentially interfering information is

an important aspect of the span task. Other studies have also suggested a link between working memory capacity and inhibitory processes (e.g. Conway and Engle, 1994).

Taking all these observations together, it seems that a simplistic interpretation of working memory span as reflecting the capacity of a central workspace is unlikely to be correct. Working memory span may involve a central workspace, but it is clearly a complex task requiring a more complex theoretical account. Such a conclusion points to the difficulty of sustaining any simple conceptualization of executive processes. Indeed, an important issue to emerge in recent studies of executive function, is whether the executive is a single, unified entity or a system that is fractionated into distinct subcomponents. This question of fractionation has led to an interest in tasks other than working memory span that capture different aspects of executive function.

### 2.3.2 Attention

The view of the executive put forward by Baddeley (1986) was substantially different from that proposed earlier by Baddeley and Hitch (1974), stemming in part from difficulties with the idea of resource trade-off. It was inspired by an imaginative attempt of Norman and Shallice (1986) to provide a unified explanation for slips of action in everyday life and the more serious disturbances of behaviour seen in patients with frontal lesions (frontal patients). One rather striking example of such a disturbance is ‘utilization behaviour’ (Lhermitte, 1983) where frontal patients show particular difficulty inhibiting stereotyped responses. For instance when a glass and then a bottle of water are merely placed in front of such a patient, the glass is picked up, filled with water and drunk. Similar behaviour is seen with other familiar objects such as a comb or a spoon.

Norman and Shallice (1986) proposed a model in which the control of cognition and action involves two levels. At the lower level is a set of learned schemata for routine sequences of actions or mental operations each of which fires automatically to a specific ‘trigger stimulus’. For example, if we overhear someone mention our own name we automatically orient our attention towards the speaker. These schemata are arranged in parallel, so that at any moment there is competition among those that potentially might fire. At the higher level sits a supervisory attentional system (SAS), a limited-capacity resource capable of intervening at the lower level. A typical example would be the SAS intervening to stop a schema from firing despite the presence of its trigger stimulus. This model explains the difficulties of frontal patients in terms of a deficit in the resources available for executive control. Thus in utilization behaviour, strongly triggered schemata fire even when they lead to contextually inappropriate behaviour. Diary studies of slips and lapses in everyday life reveal that these too often involve making an inappropriate but familiar action in a familiar context. For example, one diarist recorded intending to get his car out but as he passed through the back porch on his way to the garage he stopped to put on his Wellington boots and gardening jacket, as if to work in the garden (Reason, 1984). Such errors tended to occur when the diarists reported their attention was distracted elsewhere. The Norman and Shallice (1986) model would explain such errors in terms of distraction rendering the SAS temporarily unavailable to inhibit the strongly triggered habit of going into the garden.

Baddeley (1986) adopted the SAS as a model of executive control, thus moving away from the notion of the executive as a workspace combining both processing and storage to that of a purely attentional system. This move led more or less directly to a search for fresh ways of investigating executive processes. One such task involves generating a random stream of responses using only the digits 0–9, a surprisingly difficult task (see Box 9.6). The major source of difficulty in random generation seems to be the avoidance of stereotyped sequences such as ascending or descending series of digits, or, in the case of letters, alphabetical runs. This type of error is consistent with a theoretical analysis in which the requirement for randomness involves pitting the capacity for supervised inhibitory control against the tendency to execute strongly learned habits, sometimes called ‘pre-potent’ responses. Experimental evidence confirms that random generation is a demanding task, but shows also that it is a very complex task, suggesting that it is unlikely to be a pure measure of executive function (see Towse, 1998).

## 9.6

## Methods

### Random generation

In the random generation task, participants are asked to select items repeatedly at random from a restricted pool such as the digits 0–9 or the letters of the alphabet. Generation is usually required at a specified rate, such as one per second. Some idea of the difficulty of the task can be gained by asking someone to try it for a minute and noting down their responses. Most people soon start hesitating or repeating themselves, typically emitting stereotypical sequences such as alphabetic runs (e.g. ABC) or familiar acronyms (e.g. ITN). The degree of randomness can be estimated in various ways, one of the simplest being to count the proportion of stereotyped pairs produced. Baddeley (1986) described evidence that randomness declines systematically when either the pace of generation or the difficulty of a secondary card-sorting task was increased. These observations are consistent with the suggestion that random generation taxes a limited-capacity system.

### 2.3.3 Fractionation

In an attempt to develop the concept of the executive yet further, Baddeley (1996) proposed that the system could be fractionated into a number of separate but related functions dealing with different aspects of attention. These were focusing, dividing and switching attention, and using attention to access information in LTM. To give a general idea of these distinctions, focusing attention is required when irrelevant information has to be ignored whereas dividing is necessary when attention has to be shared between different tasks. Thus attention is focused when listening to one message and ignoring another, but divided when two messages have to be monitored simultaneously, or when different activities have to be combined, as in dual-task experiments. Attention-switching on the other hand refers to situations where attention must be repeatedly shifted from one process to another. For example, in generating a random sequence of digits, attention must constantly shift between

different retrieval plans in order to avoid stereotypical patterns of responses. This in turn is somewhat different from the role of attention within a retrieval plan when actively searching for information in LTM. Baddeley (1996) described a certain amount of empirical support for the separability of executive functions. For example, patients with Alzheimer's disease have an exaggerated difficulty in combining concurrent tasks whereas normal ageing is associated with increasing difficulty in focusing attention. However, in general the paper was theoretical and was in essence an attempt to set the agenda for future research.

One way the agenda has been taken forward is through the study of individual differences in executive function in the normal population. In one such study, Miyake *et al.* (2000) gave a large sample of students a range of tasks designed to involve different facets of attentional control. These were shifting attention, monitoring and updating information and inhibiting pre-potent responses. Analysis of the data showed that a three-factor statistical model based on these three components gave a better account of relationships among abilities than simpler (i.e. one or two-factor) models. This outcome is consistent with the general idea that executive function is fractionated, but it will be noted that the number of functions and their identity differ from Baddeley's (1996) proposal. Such a discrepancy is difficult to interpret, especially as a limitation of factor analysis is that it can only reveal the structure in the variables that are entered into the analysis. Miyake *et al.* (2000) went on to assess individual differences in a number of other tasks that are widely used as tests of executive function. The results showed that these tasks mapped onto the three putative components of executive function in different and sometimes unexpected ways. This is an interesting finding because it emphasizes the need for further development towards purer and better-understood measures of executive function.

As a general conclusion, the present state of knowledge is that executive function appears to fractionate, but it is not clear how (compare this with Chapter 2 on multiple types of attention). Thus, we still need to separate out and identify the various components of executive control. Whatever the outcome, there is a further issue of how such a diverse executive can operate in a unitary way. That is, how do the components of a many-faceted executive system interact coherently and avoid conflict in the control of perception, thought and action?

### 2.3.4 Coherence and the binding problem

It is interesting to note that the problem of coherence is not restricted to executive processes and applies to working memory more generally. Thus, if any system consists of a number of separate subsystems, then the question arises as to how the subsystems interact to ensure that the system as a whole operates in an integrated manner. For example, if visuo-spatial information about multiple objects is stored separately from verbal information about the same objects, the system must have a way of keeping track of which information refers to what object. This is sometimes referred to as the **binding problem**. Indeed, one critique of the working memory model of Baddeley and Hitch (1974) and its subsequent development by Baddeley (1986) is that by assuming separate subsystems it creates a binding problem that it fails to address (Jones, 1993). We encountered Jones' work when discussing the disruptive effect of irrelevant speech on immediate memory for verbal sequences

(Section 2.2 and Chapter 2). Salamé and Baddeley (1982) suggested that irrelevant speech enters the phonological loop, where it competes with the information to be remembered. However, Macken and Jones (1995) showed that irrelevant tones also disrupt immediate memory for verbal sequences. The amount of interference increased when the irrelevant tones or speech varied (or ‘changed state’), suggesting a common mechanism. Jones and colleagues also showed that irrelevant speech disrupts memory for spatial sequences and that, here too, variability of the unattended stimuli determines the amount of interference (Jones *et al.*, 1995). Given these observations, Jones *et al.* (1995) argued that the interference due to various types of irrelevant stimuli is best explained in terms of a common level of representation within a unitary memory system. They regarded this common ‘episodic record’ as solving the binding problem by storing combinations of features together rather than having those features dispersed over separate stores.

Do the foregoing considerations imply that the unitary view proposed by Jones is correct and that attempts to fractionate working memory should be abandoned? The answers to these two questions seem to be probably ‘not necessarily’ and ‘no’. The first answer is based on the argument that, while the similar patterns of interference across modalities suggest a common mechanism, such a mechanism could supplement rather than replace modality-specific stores. For example, the effect of variability of irrelevant stimuli might be explained in terms of the attention-grabbing property of stimulus change. Another possibility is suggested by evidence that irrelevant stimuli disrupt order information (Beaman and Jones, 1997). Thus, there might be a common serial-ordering mechanism that interacts with separate stores holding the various types of information being ordered. Perhaps the strongest reason for not abandoning fractionation is that a unitary account cannot explain the large amount of evidence for dissociations from sources other than the irrelevant sound paradigm. Nevertheless, by suggesting an alternative interpretation of the irrelevant speech effect and thereby drawing attention to the binding problem, the approach of Jones and his colleagues has made an important contribution.

In his most recent attempt to address the problem of executive control, Baddeley (2000) discusses a number of shortcomings of the tripartite 1986 model. One of these was an explicit acknowledgement that fractionation generates a binding problem. In a major revision to the model, Baddeley (2000) retained the notion of the executive as an attentional system but added to this a second component consisting of a multi-modal **episodic buffer** that integrates information across modalities and is closely associated with consciousness. This new proposal is an attempt to account for both the unitary nature of conscious experience and the coherence with which the system as a whole operates. It is too soon to evaluate the episodic buffer. For the present we note that it has much in common with Jones’ episodic record and may in part be regarded as an attempt to reconcile the tension between the two approaches of fractionation vs. integration.

## Summary of Section 2

---

- Working memory is a multi-component model, which fractionates (or partitions) cognitive activities into a series of components.
- The original fractionation was into the articulatory rehearsal loop and central executive.
- The articulatory loop is further fractionated into the phonological store and a control process of subvocal rehearsal and is now more usually termed the ‘phonological loop’.
- The central executive is an area of some ignorance, perhaps awaiting further fractionation.
- Corroborative evidence for fractionation comes from neuropsychological studies on patients with selective cognitive impairments.
- The binding problem refers to how the cognitive system keeps track of information processing about an object or task when that information is spread out over multiple independent subsystems.
- Central control needs to ensure multiple processes do not result in incoherence.
- Concepts such as episodic records (Jones) and the episodic buffer (Baddeley) attempt to solve the binding problem.
- The problem of understanding executive function in the context of working memory is actually part of a much wider field of enquiry that encompasses attention and conscious awareness.

## 3 Vocabulary acquisition

So far we have mentioned some but by no means all of the many functions of working memory and its subsystems. One that has been studied particularly closely is the role of the phonological loop in learning new vocabulary. The ability to store the sequence of phonemes making up a word must be important when encountering the word for the first time and retaining its spoken form long enough to learn it. The evidence comes from a variety of sources that include neuropsychological impairment, studies of individual differences in vocabulary size and experimental studies of word-learning.

### 3.1 Neuropsychological evidence

Some of the clearest evidence that the phonological loop must play a role in vocabulary acquisition comes from the patient, PV, whose phonological store had a reduced capacity. Although PV had a normal long-term memory for familiar items, she encountered profound difficulty in learning novel word forms. Baddeley *et al.* (1988) showed this experimentally by testing her ability to learn pairings such as *Rosa–Svieti*, where the first word was in her native Italian and the second was an unfamiliar word derived from Russian. The result was dramatic: PV showed no learning at all. However, when the members of the pairs were both Italian words she

performed normally. These observations establish a clear distinction between the processes involved in learning the two types of pairing and demonstrate a relationship between short-term phonological memory and long-term phonological learning. They also resurrect a classic debate about the relationship between short and long-term memory. Patients like PV, such as KF (see Section 1.2), who had normal LTM but extremely impaired STM, were important to the argument for separate stores. That dissociation still stands, but the fact that PV can only learn pairings of familiar items (whose phonetic structure is already stored in LTM) indicates that there is also some association between STM and LTM in the phonological domain. How should we interpret this association? One possibility is that short-term and long-term phonological memory are different aspects of the same neuroanatomical and functional system. As with Cowan's (1988) view that working memory corresponds to an activated region of long-term memory, one could think of the phonological loop as the currently active area within a phonological long-term memory system that is separate from other long-term memory systems such as semantic memory.

### 3.2 Individual differences

If learning new vocabulary items depends on the capacity to hold a phonological sequence over a short interval, then the two abilities should correlate within individuals. A number of studies have shown that children's auditory digit span correlates with their performance on tests of vocabulary (see Baddeley *et al.*, 1998). Further evidence has come from studies that assess the child's ability to repeat a nonword they have just heard (e.g. *Blonterstaping*). Nonword repetition was devised as a more demanding test of memory for phonological form than digit span, and nonword repetition is typically more highly correlated with vocabulary scores than is digit span. Of course, with a correlation it is possible the causal relationship is in the reverse direction, such that it is vocabulary knowledge that underpins the ability to repeat nonwords rather than phonological ability facilitating vocabulary acquisition. However Gathercole *et al.* (1997) found that, consistent with the latter interpretation, individual differences in the capacity of the phonological loop predict children's performance on a simulated vocabulary learning task.

As a postscript, it is interesting to note that measures of the phonological loop also correlate with vocabulary in second-language learning. Service (1992) found that Finnish children's ability to repeat English-sounding nonwords before starting to learn English predicted their English vocabulary some two years later. Moreover, Papagno and Vallar (1995) showed that polyglots selected for being fluent in at least three languages had superior auditory digit span and nonword repetition when compared with controls. The polyglots were especially good at learning word–nonword pairs but were no better than controls at learning word–word pairs.

### 3.3 Experimental studies

Yet another way of assessing the involvement of the phonological loop in new word learning is to take an experimental approach. In a series of studies, Papagno and her colleagues investigated adults learning sets of either word–nonword pairs or word–word pairs. Papagno and Vallar (1992) showed that increasing the phonemic similarity of the nonwords in a set, or the number of syllables in the nonwords,

impaired learning. However, corresponding manipulations in the word–word learning task had no effect. Papagno *et al.* (1991) found that articulatory suppression impeded the learning of word–nonword pairs but had no effect on learning word–word pairs. The absence of effects on the word–word learning task provides confirmation that the role of the loop is specific to learning novel words. These experimental differences between word–word and word–nonword learning fit well with the data on individual differences in these same tasks. However, we must bear in mind that experimental evidence that the phonological loop is necessary for learning nonwords in adults leaves open the question of whether there are stages in development when the phonological loop drives vocabulary acquisition.

## Summary of Section 3

---

- The phonological loop is involved in learning new word forms but not new associations between familiar words. These two tasks show a neuropsychological dissociation. They have also been dissociated experimentally in healthy adults.
- Individual differences in vocabulary size and vocabulary correlate with the capacity of the phonological loop in children and adults.
- However, the causal nature of the relationship between the phonological loop and vocabulary during development may be complex.

## 4 Modelling the phonological loop

Recently, a number of attempts have been made to develop mathematical and computational models of the phonological loop (Brown *et al.*, 2000; Burgess and Hitch, 1992 and 1999; Page and Norris, 1998). Part of the impetus behind these efforts is the need to explain important phenomena that the two-component model fails to address. For example, the phonological loop is only an account of immediate recall and does not say anything about learning and long-term phonological memory. Clearly, extra assumptions are needed to account for how phonological forms of newly learnt words are acquired. Even within immediate recall, the phonological loop is far from providing a complete account. Thus, an important feature of digit span and other immediate serial recall tasks is the need to remember the order of the items. Indeed, for closed sets of familiar items such as digits or letters, the most common errors are order errors. However, the phonological loop does not explain how information about order is encoded nor how order errors are generated. These omissions make a case for extending the two-component model of the phonological loop to account for long-term learning and serial ordering, while at the same time attempting to preserve its essential insights.

One argument for using modelling techniques such as computer simulation to develop and express theories is the increasing complexity of our current knowledge. As should be evident from the present discussion, one strength of the two-component account of the phonological loop is its simplicity and the ease with which

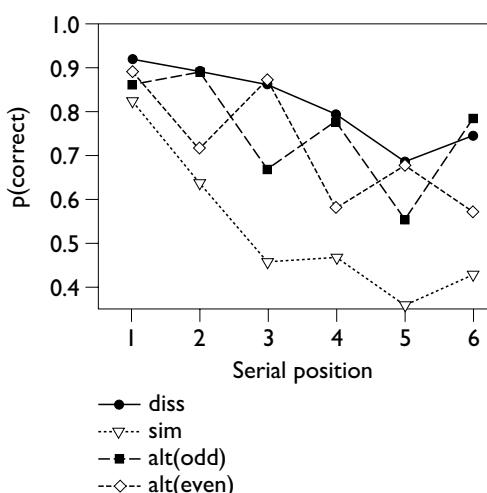
it can be used to generate testable predictions. In passing we may also note that this same strength has also allowed investigators to show where some of its assumptions are wrong (see Section 2.2.4). This is an important part of the scientific process. However, revising and extending the two-component account of the phonological loop to cope with errors and omissions runs the risk of ending up with an increasingly unwieldy theory. In particular, adding capabilities for serial ordering and learning would almost certainly render the model too unwieldy to generate clear predictions. Moving from an informal, verbal-conceptual level of theorizing to a more explicit, computational account is one way of overcoming this problem.

The most basic test of the adequacy of a computational model is whether it reproduces the same behaviour as humans when presented with the same tasks. However, this is not necessarily a very convincing test as the model-builder knows in advance the phenomena of interest and in general will have made sure the model succeeds in reproducing them. A more powerful test is to run further simulations in which the model is presented with novel experiments. The model's pattern of behaviour corresponds to its prediction about human behaviour in the same circumstances. The experiments can then be run with human participants to see whether the model's predictions are upheld. Unfortunately, it is not quite as simple as this sounds, and there are many reasons for being cautious before embarking on computational modelling. One is that developing a mechanistic account involves making extra assumptions sufficient to allow the model to 'run'. Sometimes the challenge of justifying these assumptions is hard to meet. We are fortunate in the case of auditory-verbal STM that there is a wealth of published data with which to constrain model-building. The same cannot be said, however, for executive function, and detailed computational modelling would almost certainly be premature in this case. In the following section, we describe briefly some constraints that influence the solution to the problem of how to handle serial order in the context of a detailed model of the phonological loop. Note that we do not discuss models in detail, nor evaluate their ability to explain existing experimental and neuropsychological data. Nor do we examine their ability to make novel predictions. These are all important aspects of modelling, but unfortunately there is not space to go into them here.

## 4.1 Serial order

The general problem of explaining serial order in behaviour is well known and several types of mechanism have been proposed. We will briefly describe some of these, bearing in mind that what interests us here is the specific question of what type of ordering mechanism underpins the operation of the phonological loop. According to the **chaining hypothesis**, serial order is coded by forming associations between consecutive items (e.g. Jones, 1993; Wickelgren, 1965). However, although chaining might seem highly plausible, it encounters some basic problems. One is explaining recall of a sequence containing repeated items, such as the number 2835867. If order is encoded as a chain of associations, then the repeated item (8) will be associated with not one but two following items (3 and 6). Consequently sequences containing repeated items should be difficult to recall and errors should occur after each occurrence of the repeated item. Although sequences containing repeats are more difficult to recall in verbal STM tasks, errors tend to occur on rather than after the repeated item (Jahnke, 1969). Further evidence against chaining comes

from errors in recalling sequences in which phonemically similar items alternate with phonemically dissimilar items, such as *BXDJTQVR*. These errors show a characteristic zig-zag pattern as one goes through the list, with more errors on the similar-sounding items (i.e. *BDTV*) and fewer on the dissimilar items (Baddeley, 1968; Henson *et al.*, 1996). Furthermore, dissimilar items are recalled with the same accuracy in alternating lists as in pure lists where all the items are dissimilar (see Figure 9.6). According to chaining theories, extra errors ought to occur on the dissimilar items, as these follow similar cues. For these reasons, chaining seems unlikely to explain how the phonological loop deals with serial order (though there are mathematical models that nevertheless adopt this approach, for example, Lewandowsky and Murdock, 1989).



**Figure 9.6** Serial position curves for the recall of six-item lists that varied in their phonemic composition. Diss: all items phonemically dissimilar, sim: all items phonemically similar, alt (odd)/alt (even): alternating phonemically similar and dissimilar items with similar items in odd or even-numbered serial positions

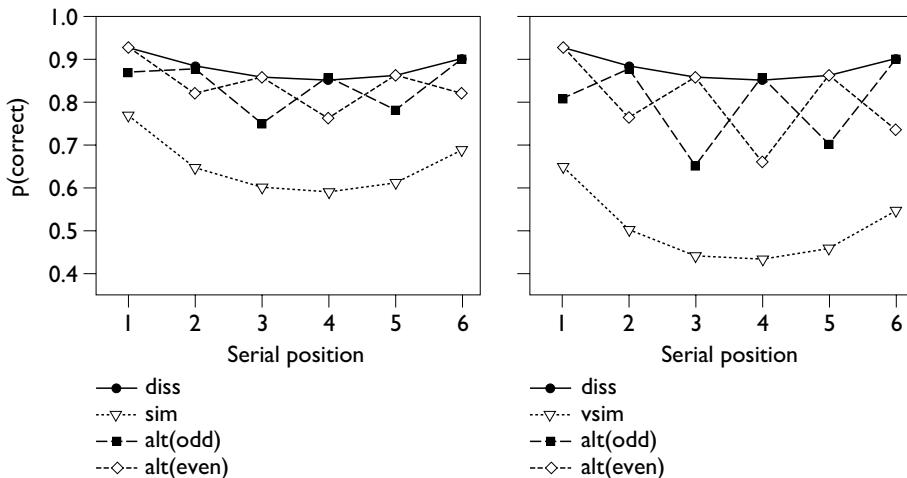
Source: Baddeley, 1968, experiment v

transposition errors decreases with their distance from the correct position (Healy, 1974). Estes (1972) proposed a mathematical model to account for this distribution of order errors, according to which positional information is encoded for each item and becomes less precise as a function of forgetting. In a related approach, recent computational models by Burgess and Hitch (1999) and Brown *et al.* (2000) propose that order is coded by associations between each item and a timing signal that varies with its position. The timing signal provides an approximate coding of position and is used to explain the distribution of order errors in a somewhat similar way to Estes. One success of this approach is that it can explain the zig-zag variation of recall with position for lists of alternating phonemically similar and dissimilar items (e.g. *BXDJTQVR*). The Burgess and Hitch model (1999) achieves this by assuming that recall of each item is a two-stage process involving first using positional information

One alternative to chaining is the positional hypothesis, according to which order is coded by associations between each item and a representation of its position within the sequence. In the simplest example of this type of model, Conrad (1965) assumed that verbal STM is composed of an ordered array of slots each containing a successive item in a list. To remember the sequence, the contents of the slots are simply read out. This simple model has no problem explaining how sequences containing repeated items are recalled. However, it cannot account for typical order errors in serial recall, where a common failure consists of an exchange between two adjacent items (e.g. recalling the sequence 318476205 as 318746205, this is known as a transposition error).

More generally the probability of

to select a candidate item and second retrieving the phonemic content of the selected item. Phonemic similarity of items is assumed to make the second of these two stages less efficient, but has no effect on the first stage. Figure 9.7 shows simulations generated by the Burgess and Hitch (1999) model. These have the same zig-zag form as the experimental data (even though the simulations do not give enough ‘primacy’, i.e. decline in recall from the start of the list).



**Figure 9.7** Serial position curves obtained by using the Burgess and Hitch (1999) model to simulate the experimental conditions of Baddeley (1968, experiment v). The left-hand panel shows simulations in which the similar items had one of two phonemes in common. The right-hand panel shows simulations in which the similar items had three of four phonemes in common (i.e. a higher degree of phonemic similarity)

A third point of view comes from non-associative models according to which encoding the order of a sequence does not involve forming novel associations. For example, in the primacy model of Page and Norris (1998), differences in the activation levels of items in memory are used to encode information about their order. Page and Norris assume that each successive item in a sequence is encoded with a lower level of activation than its predecessor. This process forms a ‘primacy gradient’ of activation levels over the list. Recall of the items in the correct serial order involves an iterative process of choosing the most strongly activated item, then the next and so on. Although this model is radically different from positional accounts, it passes the test of being able to simulate zig-zag patterns in the recall of lists of phonemically similar and dissimilar items. It is interesting to note that the model achieves this by assuming that the primacy gradient is used to select each item for recall but that a second phonological stage is used to retrieve the items phonemic composition. Thus the primacy model and the positional model of Burgess and Hitch (1999) share the idea of two stages in recall, but differ in how they assume these stages work.

So far then we can see that data on alternating lists is useful for ruling out chaining models but does not discriminate between positional and non-associative models. Fortunately, there are further data that help discriminate between these two classes of model. These relate to the **temporal grouping effect**, whereby presenting a sequence of items in rhythmic temporal groups brings about a marked reduction in

order errors in immediate recall (Ryan, 1969). Thus, recall of a sequence such as 318476205 is more accurate if it is presented as groups of three items, i.e. 318,476,205 (where the commas denote pauses). Moreover, grouping changes the pattern of order errors. Instead of the most common mistake being to transpose an item to an adjacent position, errors of recalling an item in a corresponding position in a different group become much more frequent, as in 316 478 205. These effects of grouping suggest a positional coding system in which position can be encoded at different levels. That is, a higher level codes the position of groups within a list and a lower level codes the position of items within groups. Hitch *et al.* (1996) show how their computational model captures these hierarchical effects of position in memory. Insofar as Page and Norris's (1998) primacy model encodes order on a single dimension, it cannot explain evidence for the coding of order on different levels. However, it would not be impossible to extend the model to include combinations of primacy gradients at different levels.

## Summary of Section 4

---

- One of the arguments for modelling is to go beyond the concept of a two-component phonological loop and address a wider range of phenomena such as serial ordering and nonword learning.
- Modelling is appropriate when we have a reasonably good conceptual understanding of the system we are trying to model and there are extensive data with which to constrain modelling.
- Some of the issues in modelling serial order in the phonological loop illustrate how existing data can be used to help make decisions about the underlying mechanisms.
- The chaining hypothesis can not explain certain aspects of serial order recall. However other hypotheses have had more success, for example, the positional hypothesis and non-associative accounts.

## 5 Conclusion

We have examined the concept of working memory, with particular emphasis on phonological working memory and executive control. Taking a somewhat historical approach, we have traced how the concepts of the phonological loop and the central executive emerged from previous research and how they have subsequently been developed as researchers have found out more about them. As we have seen, progress in tackling these two aspects of working memory has developed at different rates. In the case of the phonological loop, experimental evidence from a variety of sources converged on a relatively simple two-component model. In turn, this model led to insights into a range of phenomena, including the development of STM, its neuropsychological impairment and children's learning of vocabulary. Although the simple model has been shown to be inadequate in a number of details it nevertheless preserves sufficient insights to have encouraged computational modellers to develop

more detailed accounts that explain a greater range of phenomena. Such models are fairly recent and only time will tell whether this approach will prove productive.

In the case of the central executive the story is quite different. Here progress has been much slower and has consisted of various attempts to get an adequate conceptual handle on the problem. In this context the difficulty of devising satisfactory and reasonably well-understood tests of the various aspects of executive function that have been proposed should perhaps not surprise us. What is needed is a greater conceptual understanding of the various functions of executive control, one that goes beyond the promising beginning made by Norman and Shallice (1986) and develops the sorts of ideas discussed by Baddeley (1996).

In closing, we note that specifying the architecture of working memory is useful but cannot be the whole story. Thus, two important issues emerged when considering the evidence that working memory can be fractionated into a variety of subsystems. These concern how the system functions in a coherent and co-ordinated way and how it makes use of learned schemata and knowledge in LTM. Only when broader issues such as these are addressed, can we start to give a coherent account of the role of working memory in such apparently ordinary everyday activities as planning a shopping trip or reading a newspaper.

## Further reading

- Andrade, J. (ed.) (2001) *Working Memory in Perspective*, Hove, Psychology Press.  
Chapters by experienced researchers present a critical assessment of the Baddeley and Hitch (1974) model of working memory.
- Baddeley A.D. (2000) 'Is working memory still working?', *American Psychologist*, vol.56, pp.851–64. In 2001 Alan Baddeley received the American Psychological Association's Award for Distinguished Scientific Contributions. This article presents his award address, which took the form of a personal review of the current state of the Baddeley and Hitch (1974) model of working memory.
- Miyake, A. and Shah, P. (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press.  
Proponents of competing theoretical approaches to working memory were invited to present their views in a format that was designed to help clarify areas of agreement and disagreement.

## References

- Andrade, J. (2001) 'The working memory model: consensus, controversy, and future directions' in Andrade, J. (ed.) *Working Memory in Perspective*, Hove, Psychology Press.
- Atkinson, R.M. and Shiffrin, R.M. (1971) 'The control of short-term memory', *Scientific American*, vol.225, pp.82–90.
- Baddeley, A.D. (1966a) 'Short-term memory for word sequences as a function of acoustic, semantic and formal similarity', *Quarterly Journal of Experimental Psychology*, vol.18, pp.362–5.

- Baddeley, A.D. (1966b) 'The influence of acoustic and semantic similarity on long-term memory for word sequences', *Quarterly Journal of Experimental Psychology*, vol.18, pp.302–9.
- Baddeley, A.D. (1968) 'How does acoustic similarity influence short-term memory?', *Quarterly Journal of Experimental Psychology*, vol.20, pp.249–64.
- Baddeley, A.D. (1983) 'Working memory', *Philosophical Transactions of the Royal Society London*, B 302, pp.311–24.
- Baddeley, A.D. (1986) *Working Memory*, Oxford, Oxford University Press.
- Baddeley, A.D. (1996) 'Exploring the central executive', *Quarterly Journal of Experimental Psychology*, 49A, pp.5–28.
- Baddeley, A.D. (2000) 'The episodic buffer: a new component of working memory?', *Trends in Cognitive Sciences*, vol.4, pp.417–23.
- Baddeley, A., Gathercole, S. and Papagno, C. (1998) 'The phonological loop as a language learning device', *Psychological Review*, vol.105, pp.158–73.
- Baddeley, A.D. and Hitch, G.J. (1974) 'Working memory' in Bower, G. (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol.8, New York, Academic Press.
- Baddeley, A.D. and Larsen, J.D. (2003) 'The disruption of STM: a response to our commentators', *Quarterly Journal of Experimental Psychology*, 56A, pp.1301–6.
- Baddeley, A.D., Lewis, V.J. and Vallar, G. (1984) 'Exploring the articulatory loop', *Quarterly Journal of Experimental Psychology*, 36A, pp.233–52.
- Baddeley, A.D. and Lieberman, K. (1980) 'Spatial working memory' in Nickerson, R.S. (ed.) *Attention and Performance*, VIII, Hillsdale, NJ, Erlbaum.
- Baddeley, A.D., Papagno, C. and Vallar, G. (1988) 'When long-term learning depends on short-term storage', *Journal of Memory and Language*, vol.27, pp.586–96.
- Baddeley, A.D., Thomson, N. and Buchanan, M. (1975) 'Word length and the structure of short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.14, pp.575–89.
- Barrouillet, P. and Camos, V. (2001) 'Developmental increase in working memory span: resource sharing or temporal decay?', *Journal of Memory and Language*, vol.45, pp.1–20.
- Beaman, C.P. and Jones, D. (1997) 'The role of serial order in the irrelevant speech effect: tests of the changing-state hypothesis', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.23, pp.459–71.
- Brooks, D.N. and Baddeley, A.D. (1976) 'What can amnesic patients learn?', *Neuropsychologia*, vol.14, pp.111–22.
- Brown, J. (1958) 'Some tests of the decay theory of immediate memory', *Quarterly Journal of Experimental Psychology*, vol.10, pp.12–21.
- Brown, G.D. and Hulme, C. (1995) 'Modelling item length effects in memory span: no rehearsal needed?', *Journal of Memory and Language*, vol.34, pp.594–621.
- Brown, G.D., Preece, T. and Hulme, C. (2000) 'Oscillator-based memory for serial order', *Psychological Review*, vol.107, pp.127–81.

- Burgess, N. and Hitch, G.J. (1992) 'Toward a network model of the articulatory loop', *Journal of Memory and Language*, vol.31, pp.429–60.
- Burgess, N. and Hitch, G.J. (1999) 'Memory for serial order: a network model of the phonological loop and its timing', *Psychological Review*, vol.106, pp.551–81.
- Caplan, D. and Waters, G.S. (1999) 'Verbal working memory and sentence comprehension', *Behavioural and Brain Sciences*, vol.22, pp.77–126.
- Cohen, N.J. and Squire, L.R. (1980) 'Preserved learning and retention of pattern-analysing skill: dissociation of 'knowing how' and 'knowing that'', *Science*, vol.210, pp.207–9.
- Conrad, R. (1965) 'Order errors in immediate recall of sequences', *Journal of Verbal Learning and Verbal Behavior*, vol.4, pp.161–9.
- Conway, A.R.A. and Engle, R.W. (1994) 'Working memory and retrieval: a resource-dependent inhibition model', *Journal of Experimental Psychology: General*, vol.123, pp.354–73.
- Cowan, N. (1988) 'Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system', *Psychological Bulletin*, vol.96, pp.341–70.
- Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T. and Flores, L. (1992) 'The role of verbal output time in the effects of word length on immediate memory', *Journal of Memory and Language*, vol.31, pp.1–17.
- Craik, F.I.M. and Lockhart, R.S. (1972) 'Levels of processing: a framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, vol.11, pp.671–84.
- De Beni, R., Palladino, P., Pazzaglia, F. and Cornoldi, C. (1998) 'Increases in intrusion errors and working memory deficit of poor comprehenders', *Quarterly Journal of Experimental Psychology*, 51A, pp.305–20.
- Daneman, M. and Carpenter, P.A. (1980) 'Individual differences in working memory and reading', *Journal of Verbal Learning and Verbal Behavior*, vol.19, pp.450–66.
- de Renzi, E. and Nichelli, P. (1975) 'Verbal and non-verbal short term memory impairment following hemisphere damage', *Cortex*, vol.11, pp.341–53.
- Dempster, F.N. (1981) 'Memory span: sources of individual and developmental differences', *Psychological Bulletin*, vol.89, pp.63–100.
- Engle, R.W., Kane, M.J. and Tuholski, S. W. (1999a) 'Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and the functions of the prefrontal cortex' in Miyake, A. and Shah, P. (eds) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press.
- Engle, R.W., Tuholski, S.W., Laughlin, J.E. and Conway, A.R.A. (1999b) 'Working memory, short-term memory and general fluid intelligence: a latent variable approach', *Journal of Experimental Psychology: General*, vol.128, no.3, pp.309–31.
- Ericsson, K.A. and Kintsch, W. (1995) 'Long-term working memory', *Psychological Review*, vol.102, pp.211–45.

- Estes, W.K. (1972) 'An associative basis for coding and organization in memory' in Melton, A.W. and Martin, E. (eds) *Coding Processes in Human Memory*, Washington, DC, Winston.
- Gathercole, S.E., Hitch, G.J., Service, E. and Martin, A.J. (1997) 'Phonological short-term memory and new word learning in children', *Developmental Psychology*, vol.33, pp.966–79.
- Healy, A.F. (1974) 'Separating item from order information in short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.13, pp.644–55.
- Henson, R. (2001) 'Neural working memory' in Andrade, J. (ed.) *Working Memory in Perspective*, Hove, Psychology Press.
- Henson, R.N.A., Norris, D.G., Page, M.P.A. and Baddeley, A.D. (1996) 'Unchained memory: Error patterns rule out chaining models of immediate serial recall', *Quarterly Journal of Experimental Psychology*, 49A, pp.80–115.
- Hitch, G.J. (1978) 'The role of short-term working memory in mental arithmetic', *Cognitive Psychology*, vol.10, pp.302–23.
- Hitch, G.J., Brandimonte, M.A. and Walker, P. (1995) 'Two types of representation in visual memory: evidence from the effects of stimulus contrast on image combination', *Memory and Cognition*, vol.23, pp.147–54.
- Hitch, G.J., Burgess, N., Towse, J.N. and Culpin, V. (1996) 'Temporal grouping effects in immediate recall: a Working Memory analysis', *Quarterly Journal of Experimental Psychology*, 49A, pp.116–39.
- Hitch, G.J., Halliday, M.S., Dodd, A. and Littler, J.E. (1989) 'Development of rehearsal in short-term memory: differences between pictorial and spoken stimuli', *British Journal of Developmental Psychology*, vol.7, pp.347–62.
- Hitch, G.J., Halliday, M.S., Schaafstal, A. and Schraagen, J.M. (1988) 'Visual working memory in young children', *Memory and Cognition*, vol.16, pp.120–32.
- Hitch, G.J., Towse, J.N. and Hutton, U. (2001) 'What limits children's working memory span? Theoretical accounts and applications for scholastic development', *Journal of Experimental Psychology: General*, vol.130, pp.184–98.
- Hulme, C., Thomson, N., Muir, C. and Lawrence, A. (1984) 'Speech rate and the development of short-term memory span', *Journal of Experimental Child Psychology*, vol.38, pp.241–53.
- Jahnke, J.C. (1969) 'The Ranschburg effect', *Psychological Review*, vol.76, pp.592–605.
- Jones, D. (1993) 'Objects, streams and threads of auditory attention' in Baddeley, A. and Weiskrantz, L. (eds) *Attention, Awareness and Control*, Oxford, Oxford University Press.
- Jones, D., Farrand, P., Stuart, G. and Morris, N. (1995) 'Functional equivalence of verbal and spatial information in serial short-term memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.1008–18.
- Just, M.A. and Carpenter, P.A. (1992) 'A capacity theory of comprehension: individual differences in working memory', *Psychological Review*, vol.99, pp.122–49.

- Lewandowsky, S. and Murdock, B.B. Jr. (1989) ‘Memory for serial order’, *Psychological Review*, vol.96, pp.25–57.
- Lhermitte, F. (1983) “Utilization behaviour” and its relation to lesions of the frontal lobes’, *Brain*, vol.106, pp.237–55.
- Logie, R.H. (1995) *Visuo-Spatial Working Memory*, Hove, Lawrence Erlbaum Associates.
- Lovatt, P.J., Avons, S.E. and Masterson, J. (2000) ‘The word-length effect and dysllabic words’, *Quarterly Journal of Experimental Psychology*, 53A, pp.1–22.
- Macken, W.J. and Jones, D.M. (1995) ‘Functional characteristics of the ‘inner voice’ and the ‘inner ear’: single or double agency?’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.436–48.
- Miller, G.A. (1956) ‘The magical number seven plus or minus two: Some limits on our capacity for processing information’, *Psychological Review*, vol.63, pp.81–97.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H. and Howerter, A. (2000) ‘The unity and diversity of executive functions and their contributions to complex ‘frontal lobe’ tasks: a latent variable analysis’, *Cognitive Psychology*, vol.41, pp.49–100.
- Miyake, A. and Shah, P. (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press.
- Murdock, B.B. Jr. (1967) ‘Recent developments in short-term memory’, *British Journal of Psychology*, vol.58, pp.421–33.
- Nairne, J.S. (2002) ‘Remembering over the short-term: the case against the standard model’, *Annual Review of Psychology*, vol.53, pp.53–81.
- Naveh-Benjamin, M. and Ayres, T.J. (1986) ‘Digit span, reading rate, and linguistic relativity’, *Quarterly Journal of Experimental Psychology*, 38A, pp.739–52.
- Nicolson, R.S. (1981) ‘The relationship between memory span and processing speed’ in Friedman, M., Das, J.P. and O’Connor, N. (eds) *Intelligence and Learning*, New York, Plenum Press.
- Norman, D.A. and Shallice, T. (1986) ‘Attention to action: Willed and automatic control of behavior’ in Davidson, R.J., Schwartz, G.E. and Shapiro, D.E. (eds) *Consciousness and Self-Regulation: Advances in Research and Theory*, vol.4, New York, Plenum Press.
- Page, M.P.A. and Norris, D.G. (1998) ‘The primacy model: a new model of immediate serial recall’, *Psychological Review*, vol.105, pp.761–81.
- Palmer, S. (2000a) ‘Working memory: a developmental study of phonological recoding’, *Memory*, vol.8, pp.179–94.
- Palmer, S. (2000b) ‘Phonological recoding deficit in working memory of dyslexic teenagers’, *Journal of Research in Reading*, vol.23, pp.28–40.
- Papagno, C., Valentine, T. and Baddeley, A.D. (1991) ‘Phonological short-term memory and the foreign-language vocabulary learning’, *Journal of Memory and Language*, vol.30, pp.331–47.

- Papagno, C. and Vallar, G. (1992) 'Phonological short-term memory and the learning of novel words: the effect of phonological similarity and item length', *Quarterly Journal of Experimental Psychology*, 44A, pp.47–67.
- Papagno, C. and Vallar, G. (1995) 'Short-term memory and vocabulary learning in polyglots', *Quarterly Journal of Experimental Psychology*, 48A, pp.98–107.
- Paulesu, E., Frith, C.D. and Frackowiack, R.S.J. (1993) 'The neural correlates of the verbal component of working memory', *Nature*, vol.362, pp.342–4.
- Reason, J. (1984) 'Lapses of attention in everyday life' in Parasuraman, R., Davies, R. and Beatty, J. (eds) *Varieties of Attention*, Orlando, Fl., Academic Press.
- Ryan, J. (1969) 'Grouping and short-term memory: different means and patterns of grouping', *Quarterly Journal of Experimental Psychology*, vol.21, pp.137–47.
- Salamé, P. and Baddeley, A.D. (1982) 'Disruption of short-term memory by unattended speech: implications for structure of working memory', *Journal of Verbal Learning and Verbal Behavior*, vol.21, pp.150–84.
- Service, E. (1992) 'Phonology, working memory, and foreign-language learning', *Quarterly Journal of Experimental Psychology*, 45A, pp.21–50.
- Shallice, T. and Warrington, E.K. (1970) 'Independent functioning of verbal memory stores: a neuropsychological study', *Quarterly Journal of Experimental Psychology*, vol.22, pp.261–73.
- Smyth, M.M. and Waller, A. (1998) 'Movement imagery in rock climbing: patterns of interference from visual, spatial and kinaesthetic secondary tasks', *Applied Cognitive Psychology*, vol.12, pp.145–57.
- Towse, J.N. (1998) 'On random generation and the central executive of working memory', *British Journal of Psychology*, vol.89, pp.77–101.
- Towse, J.N., Hitch, G.J. and Hutton, U. (1998) 'A reevaluation of working memory capacity in children', *Journal of Memory and Language*, vol.39, no.2, pp.195–217.
- Towse, J.N., Hitch, G.J. and Hutton, U. (2000) 'On the interpretation of working memory span in adults', *Memory and Cognition*, vol.28, no.3, pp.341–8.
- Turner, M.L. and Engle, R.W. (1989) 'Is working memory capacity task dependent?', *Journal of Memory and Language*, vol.28, pp.127–54.
- Vallar, G. and Baddeley, A.D. (1984) 'Fractionation of working memory: Neuropsychological evidence for a phonological short-term store', *Journal of Verbal Learning and Verbal Behavior*, vol.23, pp.151–61.
- Wickelgren, W.A. (1965) 'Short-term memory for repeated and non-repeated items', *Quarterly Journal of Experimental Psychology*, vol.17, pp.14–25.



---

# PART 4

# THINKING

---

Introduction

**Chapter 10 Problem solving**

*Alison J K Green and Ken Gilhooly*

**Chapter 11 Judgment and decision making**

*Peter Ayton*

**Chapter 12 Reasoning**

*Mike Oaksford*

# Introduction

In Part 4, the focus shifts to what have been termed thinking processes. Specifically, the chapters address three distinct kinds of thinking that arise in different kinds of task – in general terms, these are tasks that require us to solve problems (Chapter 10), come to judgments and make decisions (Chapter 11) and to reason and draw conclusions (Chapter 12).

In Chapter 10, Alison Green and Ken Gilhooly address human problem solving. You might think that problem solving is a somewhat artificial activity, inspired by abstract and contrived problems such as crossword puzzles, or the Rubiks cube. But problem solving in cognitive psychology is intended to encompass a wide range of activities in which we need to identify the solution to a current problem. Everyday problems range from the easy, such as how to make a cup of tea in someone else's kitchen, to the complex, such as how to achieve career success. Everyday problems are not always easily defined – think of the different problems that need solving in order to achieve a successful career – and so psychologists have often relied on more formally specified problems, of which the authors of Chapter 10 provide numerous examples.

In Chapter 1 it was pointed out that cognitive psychology tends to avoid the study of individual differences because it aims to understand cognitive processes in general. However, in Chapter 9 of Part 3, we saw how the study of individual differences can help us to evaluate theories of cognition and, conversely, how the theories can help us to understand the nature of the individual differences. In Chapter 10, the importance of individual differences is emphasised again. Firstly, the authors note that some individuals are novices in solving certain classes of problem and some are expert. In playing chess, for example, some individuals achieve grand master status whereas others, while knowing the rules of the game, are considerably less skilled. The problem-solving approaches of these groups differ in important ways, and so it is not true to say that people in general tackle chess problems in the same way. Secondly, as the authors also point out, experts themselves differ from one another in relevant ways – that is, they do not form a homogeneous group – and novices also differ from one another.

Certain aspects of the study of problem solving that are raised in Chapter 10 become themes for the whole of Part 4. One is that cognitive psychologists place at least as much emphasis on the study of errors in problem solving as they do on cases of success. Indeed, as we shall see, errors provide important information concerning underlying cognitive processes, and this theme is continued in Chapters 11 and 12.

Another theme established in Chapter 10 is the importance of establishing a framework within which phenomena can be analysed and understood, and which in turn can be used to derive new research questions. In Chapter 10, Alison Green and Ken Gilhooly introduce the notion of a ‘state–space’ and show how this notion can help us to understand problems and to analyse human performance when attempting to solve them.

In Chapter 11, Peter Ayton introduces the topics of judgment and decision making. How do we form judgments and make decisions, and how best should we understand and analyse these cognitive activities? One thing the author makes clear from the outset is that it is possible to develop different kinds of theory, depending on

the starting point and purpose of the theorist. A theory could, for example, emphasise how judgment and decision-making processes *ought* to proceed, and use this as a basis for analysing human performance. Alternatively, a theorist could take as their starting point an understanding of how people *actually* make decisions, including poor ones. That there are these two approaches – normative and descriptive – continues the theme, established in Chapter 10, of the importance of researchers establishing an appropriate explanatory framework. In essence, the chapter can be seen as an extended discussion of whether two particular normative approaches provide an adequate understanding of human judgment and decision making.

In the first part of Chapter 11, Peter Ayton discusses the normative theory of subjective expected utility, and its use as a vehicle for understanding human decision making. As you will see, this theory requires us to express the likelihood of particular outcomes as mathematical probabilities, though you may be relieved to know that the key mathematical ideas are relatively simple. Probabilities are also required to understand a normative approach to judgment under uncertainty, which involves the application of Bayes' Theorem.

While the normative approaches rely on mathematical formulations of how judgment and decision making ought to proceed, the majority of the chapter discusses evidence that human performance actually departs from these mathematical standards – that human performance is characterised by apparent errors. The chapter compares such normative theories with descriptive accounts of actual human performance – prospect theory and the heuristics and biases approach. Although these accounts are seen ultimately to be more successful, the normative approaches nonetheless play an important role in helping researchers to develop a more appropriate explanatory framework. In particular, observations that human performance deviates systematically from mathematical standards have provided researchers with extremely valuable information.

The particular descriptive approaches discussed in the chapter are not without their problems however. One important line of criticism comes from an approach that considers the adaptive function, in evolutionary terms, of decision-making processes. Some researchers have argued that such processes would have evolved to be 'fast and frugal', and the chapter cites evidence in favour of this view.

Some of these concerns can also be found in Chapter 12. In this chapter, Mike Oaksford shows how research into reasoning also started from a formal framework for understanding how people ought to reason – that of logic. The first part of Chapter 12 outlines the nature of logic, and in particular the forms of reasoning that are taken to be logically valid, and those taken to be logically invalid. As with the normative approaches discussed in Chapter 11, logic provides a valuable framework for trying to understand human reasoning. It has also helped researchers to establish the core phenomena of reasoning, and to develop paradigms to investigate these.

As with Chapter 11, a key observation also running throughout Chapter 12 is that human reasoning systematically departs from the normative standards established by logic. Predicting and understanding these logical errors has thus become an important benchmark for theories of human reasoning. A number of such theories have been developed, and the chapter focuses mostly on three approaches – mental logic, mental models and the probabilistic approach. Each of these approaches is

evaluated against evidence that has accrued from the use of two key paradigms or tasks – conditional inference, and the Wason selection task – and we see again the use of expected utilities and also Bayes’ theorem.

Specifically in connection with a version of the Wason selection task, the chapter also discusses a fourth approach to reasoning. This approach emphasises the importance of theories of reasoning positing processes that can be seen to have an evolutionarily adaptive function. Mike Oaksford also discusses the relationship between logical reasoning and IQ, showing once again the importance of individual differences.

As suggested above, there are a number of themes running throughout the three chapters of this Part 4. One theme that has not been mentioned so far is rationality. Errors, or departures from a logical or mathematical standard, could be taken as signs of the intrinsic irrationality of human thought. After all, so the argument goes, if human beings are rational they ought to solve problems, make judgments and decisions, and reason according to certain standards, often assumed to be provided by formal models, such as mathematics and logic. Departures from these mathematical and logical standards then would be signs of irrationality. However, although these three chapters do not take rationality as their central focus, the establishment of theories of actual human performance provide grounds for understanding rationality differently. In connection with Chapter 11, for example, perhaps it would be rational to use heuristics and biases to come to a judgment, or, related to Chapter 12, perhaps it would be rational to rely on a probabilistic method for tackling a task, even if this sometimes generates logical ‘errors’.

Finally, one thing you may notice about the three chapters in this part is that there is very little mention of computer modelling, neuropsychology or neuroimaging. Although this inevitably reflects to a degree the practical limits imposed by writing a chapter to a certain length, it also reflects a particular emphasis common to the three chapters with trying to develop an appropriate explanatory framework. If you recall the discussion of Marr’s levels of explanation in Chapter 1, it is as if cognitive psychologists studying thinking are still trying to establish what is computed when we reason, or make decisions, or solve problems. Consistent with this is the fact that some researchers are appealing to evolution to help provide an explanatory framework. The suggestion that researchers are still grappling with difficult questions at Marr’s computational level, provides one way of understanding the emphasis on formal approaches – such as state–spaces in Chapter 10, probability in Chapter 11 and logic in Chapter 12. Such formal approaches provide an idealised model of what needs to be computed, i.e. Marr’s level 1 (idealised because we observe systematic departures from this in actual human thinking). The combination of these models and observations of systematic human ‘error’ provides researchers with an effective means for analysing human thinking. As research develops further, and detailed questions subsequently arise concerning the actual processes by which thinking is achieved, neuropsychological, neuroimaging and computer modelling work is likely to become much more relevant.

Alison J K Green and Ken Gilhooly

---

## 1 Introduction

Problem solving is an essential, familiar and pervasive part of everyday life. Examples are all around us. Consider an infant trying to fit shapes into the appropriate holes of a shape-sorting toy, or a child trying to count out the correct sum of money to buy a new music CD, or perhaps an adult weighing up the pros and cons of a job offer. While we shall be looking here at examples of human problem solving, problem solving occurs in animal life too. Naturally occurring instances include tool use and searching for food. Problem solving in all its manifestations is an activity that structures everyday life in meaningful ways.

By studying the myriad ways in which we solve problems, we hope to learn how problems are solved effectively, and to understand what goes wrong when they are not. Why should we be interested in finding out about *unsuccessful* problem solving? An interesting aspect of failure involves investigating the errors that people make, in order to understand why a particular error occurred and to try to prevent it from happening again. Some errors are made with little cost (for example, sprinkling coffee instead of sugar over cornflakes at breakfast), but other errors can be quite catastrophic (for example, an oil-laden vessel running aground near a shoreline community of wildlife). Diagnosing errors and re-designing tasks to guard against critical errors are important applications for problem-solving research. We discuss others later on in the chapter.

Where does problem solving ‘sit’ in relation to other areas of psychology? Problem solving is an activity that draws together the various different components of cognition. For instance, linguistic skills are used to read about a political problem and engage in a debate about it. Visual perception is necessary for understanding a graphically presented engineering problem and for drawing a solution. We use memory to recover any prior knowledge we might have that could be relevant to solving a new problem, and attention plays a role in all problem solving.

Problem solving takes place over time, interleaving a range of cognitive processes and drawing upon pieces of knowledge, which are represented in various ways. The notion of ‘representation’ is central to cognitive psychology, as you will see later in Chapter 17. For now, we shall ask you to assume that information used in problem solving comes to be internally represented.

Because problem solving occurs over time, we need to study not just the cognitive processes and mental representations involved in problem solving, but also the ways in which these processes and representations interact with others. Problem solving then, like reasoning, judgement and decision making, is an activity that necessarily draws upon a range of cognitive processes.

In fact, problem solving often involves reasoning, judgement and decision making. For instance, a general practitioner gathering information about a sick patient’s symptoms may deduce from the description the patient gives that the problem is a bacterial, rather than a viral, infection. The doctor may then make a

judgement about the severity of the infection, before making a decision on an antibiotic to prescribe.

In this chapter, we examine the ways in which individuals approach a variety of problem types, ranging from simple, puzzle problems to more complex, real-world problems. Everyday problems can be complex and challenging, with constraints in operation that mean that the solution we choose or find may not be an ideal one. As individuals we can, if we are reasonably adept, persistent or just plain lucky, solve many of the problems that come our way. Quite often though, our initial attempts fail and we have to turn to another source to help solve the problem – a manual, for instance, in the case of a tricky computer installation problem, or perhaps someone knowledgeable in the problem area if all else fails.

While some problems may be viewed as unwelcome obstacles, to be avoided where possible, there are occasions where we keenly seek out problems to occupy our time. An expert mathematician, for instance, may spend hours identifying a problem, primarily for the pleasure derived in exploring and solving it. Similarly, ‘make-over’ television programmes can be very entertaining as viewers watch an undecorated room or a derelict garden transformed by the experts in a matter of days into something quite different. Some of us undoubtedly while away the hours working on tricky crossword puzzles, computer games or trying to make (or repair) something at home.

Our aim in this chapter is to present an overview of research on problem solving. In doing this, we have had to be selective, and have elected to present work that we believe has been both influential and interesting to try to give you a flavour of what has been going on in the field. As you read on, you will learn that how people represent problems is a principal determinant of problem-solving success. Much of the research we shall examine addresses the question of which factors influence the construction of a problem representation.

We aim to show you how ideas about problem solving have developed and changed. You will learn that early work on problem solving was often confined to puzzle problems and that, later on, researchers became interested in more complex domains, where knowledge and experience are central to successful problem solving. The issues we shall explore centre on the nature of problem solving, and the relationships between problem solving, learning, experience and creativity. The kinds of questions we shall be asking include:

- 1 What are the different forms of problem-solving activity?
- 2 How do we solve different sorts of problems?
- 3 Why is representation important?

First, we shall try to define a ‘problem’ and then look at conceptions of problems and problem solving.

## 1.1 What is a ‘problem’?

Before reading on, try the following activity.

### ACTIVITY 10.1

---

What do you think are the defining attributes of problems? You will probably draw upon some examples from your own experience to help you. You might like to think back to Chapter 5 to help you think of problems in terms of properties, categories and so on. Make a list of all the attributes you can think of, and then try to construct a sentence or two, defining problems. Try not to spend more than a couple of minutes on this.

### COMMENT

The answer to the question, ‘What is a problem?’ is not at all easy, as the exercise shows. You may find that you want to vary your definition, depending upon the type of problem you have in mind. You may find that you cannot come up with a definition at all, or that you came up with several and cannot choose between them. Of course, if it is difficult to define what we mean by a ‘problem’, then it becomes even more difficult to construct models and theories of problem-solving behaviour, and to compare and contrast such models and theories. Clear definitions are therefore important at the outset.

---

Consider the following examples of problems that a given individual might come across, some more commonplace than others:

- 1 Who can I ask to babysit the children so that I can go out next Thursday evening?
- 2 How can I make sure that the stone I have just played in my game of Go<sup>1</sup> ‘lives’?
- 3 Is there a way I can arrange some paper pattern pieces on my dress material so that all the pieces fit and I don’t have to buy any more material?

The problem in the first example is finding a babysitter, which could involve searching through an address book, recovering some names from memory or calling round on a friend and asking for a favour. If these fail to produce a name, then other options include carrying out a more extensive search. A bit of inspired guesswork might lead to an internet babysitting site, and locating a babysitter to solve the problem. Notice that there are several ways to satisfactorily solve this problem, and that the possible solutions vary in degree of novelty. The availability of a possible solution method may well vary too, depending upon the context (is there time to explore different possible solutions to the problem?), social setting (is

---

<sup>1</sup> The aim in Go is to use stones (one player takes black stones, the other white) to surround territory on a board. A stone (or stones) ‘lives’ if it cannot be surrounded, and therefore removed, from the board. Territory is ‘won’ if stones of one colour completely surround stones of the other colour, and the winner is the player who surrounds the most territory.

there a network of likely babysitters to call upon?) and culture (is it acceptable to use an internet babysitting agency?).

The problem in the second example centres on the ancient Korean game of Go. Here, the problem seems more to do with experience, knowledge and skill, although motivation (does playing badly matter?), personality and emotion (are there personal costs in playing badly?) and cultural factors (different cultures have different conventions for Go) may well be involved too. Again, there are different solutions available, in that a number of different moves may achieve the goal of ensuring the ‘survival’ of the stone in question.

The final problem is different again, because it involves perception in ‘seeing’ how to lay all the pieces out, together with some creative or lateral thinking in optimizing layout so that all the pieces do indeed fit correctly. There may be one or more possible ways to arrange the pieces and solve the problem, one of which may be better (for example, in ensuring that cut pieces of fabric fit together in a way that matches up a pattern at seams).

These examples show that while problems do share some common characteristics (see the discussion of concepts in Chapter 5), it is also true that different problems are affected by different factors, both internal (for example, motivation and personality) and external (for example, social and cultural factors).

Duncker (1945, p.1) offered a concise definition of a problem that captures something of the essence of our everyday experience of problems. He wrote that: ‘a problem exists when a living organism has a goal but does not know how this goal is to be reached’. The definition is still serviceable today because it conveys the notion of a ‘gap’ between a current state and a goal or desired state. If there are no obstacles preventing the individual from moving from the current state to the desired state, then there cannot be said to be a problem. Problems, then, consist of three components: a starting state, a goal state and a set of available actions to move from the starting state to the goal. According to this type of definition, what constitutes a problem for one individual may not be a problem for another. For instance, a moderate Go player might have some difficulty ensuring that a newly placed stone survives in her current game if her opponent is a much stronger player than her. The stronger opponent however will almost certainly have considerably less difficulty in making his stones live.

So far, we have tried to present some defining characteristics of problems. Before we move on to discuss research on problem solving, we want to draw your attention to one of the principal methods used in problem-solving research: protocol analysis.

## 1.2 Protocol analysis in problem-solving research

Cognitive scientists make extensive use of a method known as ‘protocol analysis’. At the core of the approach is the view that information represented in working memory may be verbalized, either directly if in verbal form, or through transformation if in non-verbal form. Information retained in long-term memory must first be transferred to working memory before it can be reported. Thus, the ‘protocol’ of protocol analysis is a verbal account of information that is heeded as a task is carried out. The protocol that results from thinking aloud is assumed to preserve the order in which information has been heeded. Using careful instructions,

and with a little practice, most people can ‘think aloud’, either while working on a task, or immediately after completing a task.

Protocol analysis depends upon fundamental assumptions, the most basic of which are that cognition is information processing, that information is stored in different memory stores, and that recently acquired information is retained in working memory.

The method has many uses, particularly in helping to identify differences between individuals in terms of information heeded, and processes and strategies used, as a task is carried out. Let us suppose that our research question centres on investigating cognitive processes in arithmetic, and that we have asked two individuals to think aloud while calculating the sum of  $63 + 37$ . Both give the answer ‘100’. Did both arrive at the answer in the same way? One way of addressing this question is to compare the verbal reports produced:

First individual: ‘OK, what is the sum of 63 plus 37? Easy – that’s 100’.

Second individual: ‘What is the sum of 63 plus 37? 60 plus 30 ... 60, 70, 80, 90. 3 plus 7 is 10. 90 plus 10 is 100. It’s 100.’

The first individual simply reads out the problem statement and then reports the answer. There is little evidence of any problem solving here, and the answer appears to be readily available – it is as if the individual is retrieving a number fact. The second individual also begins by reading the problem statement, but then goes about the problem rather differently. The protocol suggests that her strategy is a ‘counting on’ strategy, starting with 60, then counting on 30, giving 90. She then adds the units 3 and 7, giving 10, and finally adds 10 to 90 to give the answer. The example shows that different people can arrive at the same answer, but use different methods. It also shows that protocol analysis can reveal useful information about strategies underlying behaviour.

Protocol analysis is a very useful tool for identifying different strategies people use in problem solving – strategies that may not be obvious from problem solutions alone. (In Chapter 16, you will encounter some models of the ways in which simple arithmetic problems, like those above, may be solved.) Of course, there are situations where protocol analysis is not a suitable approach (for example, where the requirement to think aloud might actually change the way in which the task is carried out).

It is important to recognize that thinking aloud is not the direct externalization of our cognitive processes. Rather, mental processes may be inferred through the careful analysis of verbal protocols. We illustrate an application of protocol analysis in Box 10.1.

**10.1****Methods****Protocol analysis applied to medical diagnosis**

Medical diagnosis is a complex skill, requiring the clinician to bring to bear his or her knowledge and skill in accurately diagnosing a given patient's disorder. Expert clinicians have acquired both biomedical and clinical knowledge. Biomedical knowledge includes knowledge of anatomy, biochemistry and physiology, while clinical knowledge is often expressed in terms of associations between symptoms, or clinical findings, and disease categories. There has been some debate over the extent to which expert clinicians use biomedical knowledge in making diagnoses. Lesgold *et al.* (1998) found that expert clinicians made extensive use of biomedical knowledge, whereas Boshuizen and Schmidt (1992) found they made very little use.

Gilhooly *et al.* (1997) hypothesized that when experts can use contextual information (e.g. patient's age, gender and lifestyle habits) to aid a diagnosis, their use of biomedical knowledge may be suppressed. Gilhooly *et al.* tested this hypothesis through their analysis of think-aloud protocols produced by clinicians varying in skill level. They asked a group of clinicians to interpret electrocardiogram (ECG) trace information. The ECG is regularly used to assess the electrical activity of the heart, and to help identify abnormal patterns of activity that might indicate an underlying problem. Skill is required to interpret an ECG trace, and to use this in making an accurate diagnosis, which then becomes the basis for a patient's treatment regime.

Gilhooly *et al.* asked groups of registrars (the 'experts'), house officers (the 'intermediates') and third-year medical students (the 'novices') to think aloud while they studied and diagnosed eight different ECG traces, presented with no context information. They then analysed the protocols, examining them for evidence of biomedical and clinical knowledge. For example, use of key terms such as 'polarization', 'activation' or 'conducting', were categorized as biomedical references. Use of words such as 'chronic' or 'hypertension' were classified as clinical references. Clinicians also described the ECG traces directly in their protocols, giving a third category of words. In this way, the protocols produced by the clinicians were segmented into much smaller chunks, corresponding to clinical or biomedical inferences, or trace descriptions.

Reassuringly, the more experienced and skilled the clinicians were, the more accurate their diagnoses. The results of the protocol analysis showed that more skilled clinicians made more extensive use of their biomedical knowledge than less skilled clinicians, particularly in evaluating possible diagnoses. They also made more use of their clinical knowledge than the less skilled clinicians.

What does this study tell us? First, it resolves the apparently discrepant findings in the literature. Increased use of biomedical knowledge is associated with expertise, when clinicians are not able to use shortcuts to aid a diagnosis. Second, the study shows that protocol analysis can be a very useful tool in helping us to understand problem solving in real-world situations. Verbal protocols can give valuable insights into knowledge and processes involved in problem solving.

# Summary of Section 1

---

- Problems involve a start state, a goal state and a set of actions or operators that may be applied to move from one state to the next until the goal is achieved.
- Protocol analysis is a key method in problem-solving research.

## 2 ‘Simple’ problem solving

In this section we discuss themes and issues in research on what might loosely be termed ‘simple’ problem solving, although as you shall see, the problems used are not always simple to solve. So-called ‘simple’ problems, which do not require extensive background knowledge, are sometimes known as ‘puzzles’ and have often been used in research as most participants can attempt such problems within a reasonably short time. The issue of representation, and the various ways in which manipulations of problems affect representation, and in turn, problem-solving performance, is very much at the centre of this branch of problem-solving research.

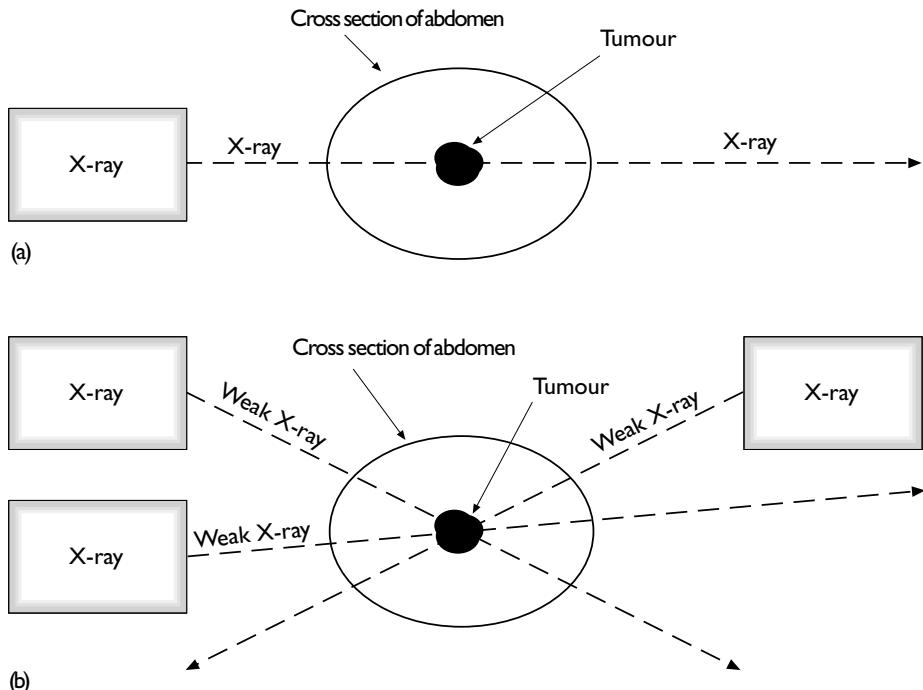
### 2.1 The Gestalt legacy

Simple problem solving began to be studied intensively from the 1910s by a group of German psychologists known as the **Gestaltists**. The hallmarks of the Gestalt approach were the phenomenon of insight, and the view that the whole is greater than the sum of its parts. Insight has famously been labelled the ‘aha!’ phenomenon, in that sudden restructuring or re-representations of a problem can sometimes lead to a solution.

The Gestalt school particularly emphasized the role of **insight** in problem solving. An example can be found in the story of young Gauss (Hall, 1970) who later went on to become a prominent mathematician (well known for deriving the formula for the normal distribution curve). As a young schoolchild, Gauss surprised his teacher by very quickly producing the correct answer to the sum of all the numbers from 1 to 100. He gave the answer (5050) not by very fast mental arithmetic but by noticing a pattern in the number sequence, viz., that the numbers form pairs ( $1+100=101$ ,  $2+99=101$ ,  $3+98=101$ ... and so on). There are 50 pairs and each pair sums to 101 hence the answer is 5050. In this example then a good structuring, or representation, of the problem, helps considerably.

The processes of restructuring were investigated further by Duncker (1945) who asked participants to think aloud as they tackled problems that required insight to solve. An example is the X-ray problem (see Figure 10.1 below). Participants were shown a diagram and told that it represented a patient with a tumour in the centre of his body. The problem was how to use an X-ray apparatus to destroy the tumour without destroying the surrounding healthy tissue. Participants usually tried alternative restructurings of the problem in terms of sub-goals that could lead to solutions. Thus, the major goal could be achieved if a sub-goal of avoiding damage to healthy tissue could be achieved. The most common solutions involved a sub-goal of lowering intensity of rays on their way through the healthy tissue. This sub-goal

led to the solution of using a number of weaker rays, which then converged on the tumour at lethal intensity, thereby destroying the tumour. (An alternative solution involved using a lens to focus a broad band of weak rays on the tumour so that lethal intensity was reached only at the focal point.)



**Figure 10.1** The X-ray problem

The Gestalt psychologists also investigated cases where insight was generally not achieved because participants were trapped by misleading representations that prevented solution. So-called ‘set’ effects arise when learned or habitual ways of tackling a problem prevent the solver from identifying better and simpler methods, or when unwarranted assumptions are made.

Set can be induced by experience with a series of similar problems. Luchins and Luchins (1959) studied problem sets in a series of experiments using water jar problems (presented as a pencil-and-paper exercise). In these tasks participants were asked to say how one could get exactly a specified amount of water using jars of fixed capacity and an unlimited source of water. For example:

Given three jars (A, B, and C) of capacities 18, 43 and 10 units respectively, how could you obtain exactly 5 units of water?

The solution may be expressed as B-A-2C. After a series of problems with that same general solution, participants had great difficulty with the following problem:

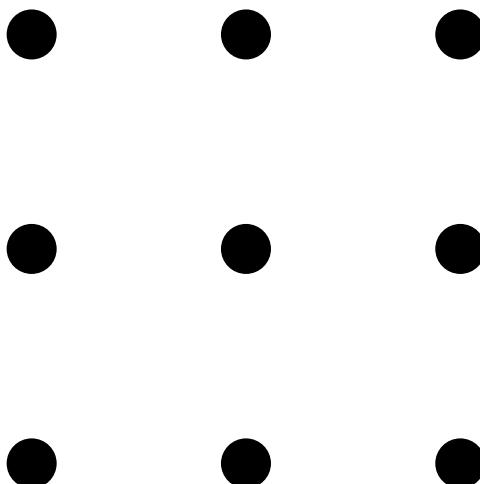
Given three jars (A, B, and C) of capacities 28, 76 and 3 units respectively, how could you obtain exactly 25 units of water?

In fact, the solution to this problem is quite simple (i.e. A-C) but when this problem is presented after a series of problems involving the long solution (B-A-2C) many participants used the inefficient method and either failed to solve the problem, or took considerably longer to use the A-C method than did a control group of participants.

Figure 10.2 illustrates the 9-dot problem, often used to investigate this particular type of set effect. This problem is another example of the set effect, this time produced by the *layout* of the task. Try the following activity.

### ACTIVITY 10.2

---



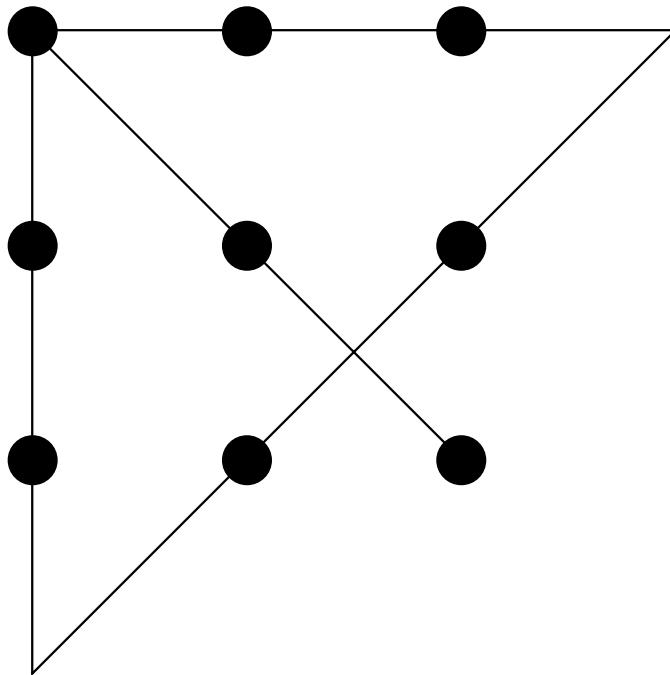
**Figure 10.2** The 9-dot problem

Starting from any point, draw four straight lines (without lifting the pen from the page) so that each of the nine dots has at least one line running through it.

#### COMMENT

Most participants interpret the instructions as meaning that they must stay within the square shape of the dots; however, a solution is not possible without breaking this set and going outside the square. Figure 10.3 shows a solution.

A related block to effective problem solving, known as '**functional fixity**' (also identified by work in the Gestalt tradition) tends to be observed when an object has to be used in a *new way*. Duncker (1945) carried out the classic study of functional fixity using the 'box' (or 'candle') problem. In this task, participants were presented with tacks, matches, three small boxes and three candles. The problem was to mount the candles side by side on a door, so that they could burn safely. For one group of participants the boxes were empty but for the other group (experimental group) the boxes were used as containers and held matches, tacks and candles. The solution is to use the boxes not as containers but as platforms and fix them to the door using the tacks. It was found that the solving rate was much higher in the control group than in the experimental group. Duncker explained this result in terms of a failure to



**Figure 10.3** A solution to the 9-dot problem

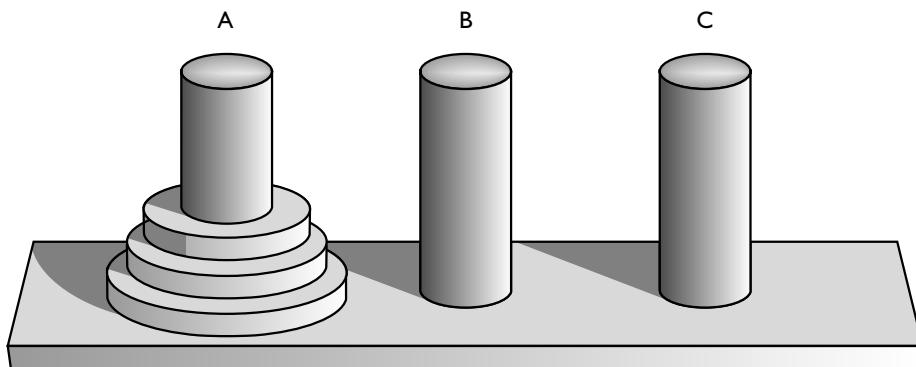
perceive the possible platform function of the boxes when they were presented as containers. Functional fixity has been independently demonstrated and explored further in a number of later studies (e.g. Adamson and Taylor, 1954; Glucksberg and Danks, 1968). The phenomenon appears to be a robust one and is a likely source of difficulty in real-life problem solving.

There are many real-world examples of functional fixity effects. An interesting early example is the history of the steam engine. In 1775, the first Watt steam engine was used to pump water from a colliery, thus solving the problem of flooding. Before steam engines, either buckets or inefficient suction pumps had been used. It was some years before it was appreciated that steam engines could be used for locomotion as well as pumping water.

These early studies demonstrate the importance of representation and its impact upon problem solving. Later research, as we shall see in the next section, went on to examine representational effects in a wider range of problems in more depth.

## 2.2 Representation in puzzle problem solving

The ‘representational effect’ has been acknowledged for some time in problem-solving research. Simon and Hayes (1976) constructed several versions of the Tower of Hanoi problem (see Figure 10.4 below), which involves discs of varying sizes arranged on three pegs. The goal is to move the three discs from one peg (e.g. peg A) to another peg (e.g. peg C) using a sequence of legal moves. Typical constraints for this problem are that a larger disc can never be placed on top of a smaller disc (see below though, for a variation on this rule) and only one disc may be moved at a time.



**Figure 10.4** The three-disc version of the Tower of Hanoi problem

‘State–space’ diagrams present a given problem’s state at each move juncture. Problem structures can be mapped out and compared using state–space diagrams. (The state–space diagram for the Tower of Hanoi is quite complex. See Figure 10.5, though, for another example of a state–space diagram, drawn to illustrate the structure of the simpler ‘hobbits and orcs’ problem, previously known as the ‘missionaries and cannibals’ problem).

Problems that share the same underlying structure (i.e. have identical state–space diagrams) are said to be **isomorphic**. Simon and Hayes contrasted two structurally similar versions of a ‘monster’ problem. Their ‘monster’ problem was itself isomorphic to the Tower of Hanoi problem. In the ‘move’ version, differently sized monsters transferred globes of different sizes to each other according to a set of rules. In the ‘change’ version, monsters differing in size each held a globe, which had to be changed in size to conform to particular rules. Despite being isomorphs, the ‘change’ problem was considerably harder to solve than the ‘move’ problem. People seemed to construct rather different representations of the two problems; the representation constructed for the ‘move’ problem entailed simpler processing operations than that constructed for the ‘change’ problem.

Zhang and Norman (1994) developed a theory to account for representational effects with these sorts of problems. They designed a number of isomorphic versions of the basic Tower of Hanoi problem, and explored ways in which different rules influenced problem difficulty. Their theoretical framework distinguishes between **‘internal’** and **‘external’** representations. Internal problem representations entail a processing and representational burden, because the information needed to solve the problem has to be encoded and maintained in some form. Internal rules then are rules that need to be memorized, such as:

- 1 Only one disc may be transferred at a time.
- 2 A smaller disc may never be placed on top of a larger disc (notice that this is the reverse of the usual rule for this problem).

External rules differ, however, in that they are not stated explicitly in the instructions, but are implied or necessitated by the problem itself. For instance, a form of the Tower of Hanoi where discs are replaced by cups of different sizes,

filled with coffee, involves an external version of Rule 2 above (a smaller cup would fall into the larger cup, spilling coffee). The environment then can provide constraints, so some rules need not be internalized. Size and location are properties that need not be internally represented, since differences in size and location are readily perceived. However, dimensional information may be represented internally. For instance, if colour is used to represent some task-relevant information, then the relationships between colours and information may have to be learned.

External representations appear to make problem solving easier, although they also change the nature of the task. We return to this point later in Section 5 in our examination of the relationship between problem solving and learning.

Some problem representations have attributes that may hamper (or facilitate) problem solving. Once a problem has been encoded and represented, problem solving may be described as a search through a set of possible moves (or ‘problem space’).

### 2.3 The information processing approach: problem solving as search

Solving a problem may require us to find a suitable sequence of actions drawn from a small set of actions (for instance, moving a series of coloured tiles around a small board until they form a particular pattern or picture). Alternatively, solving some problems may entail selecting or discovering a single action from a large set, for example, using one object that can meet the goal from all objects known to the solver. Within the information processing approach, problem solving is generally seen as a **search** process.

The initiator of problem solving is a current goal, that is, a representation of a state that is desired but not currently true. Therefore, goals direct the course of thinking by guiding retrieval of goal-relevant material and aiding in the assessment of directions of search as promising or not. Search may proceed in a **forwards** direction from the starting state by generating possible actions, evaluating the results of those actions and then choosing for further exploration those with best outcomes when assessed against the goal. Search may also proceed **backwards** from the goal by using a **problem-reduction** or **means–ends** approach, which breaks down the overall goal into sub-goals that should be easier to achieve.

For example, the problem of booking a trip to New York from London might be broken down (or ‘reduced’) into sub-goals that could include ‘buying a ticket’ and ‘getting our passports updated’. ‘Buying a ticket’ can be broken down further into sub-goals such as ‘finding an airline’ and ‘deciding on travel dates’ (which in turn can be broken down further still to ‘check availability of seats’). The initial goal of getting to New York cannot be achieved until all the steps, and their required conditions, have been identified. The problem is solved by working ‘backwards’ from the goal, starting with the first sub-goal for which the conditions may be met (e.g. establishing that there are seats available for the date on which we want to travel and selecting these), and then completing other sub-goals until the major goal is achieved and the trip is booked.

A number of studies of search in problem solving have used the Tower of London task (which is similar to the Tower of Hanoi). The problem has a number of variants but basically requires participants to first plan out how to move a set of coloured same-sized discs arranged over three pegs from a starting pattern to a target pattern by moving only one disc at a time. Search processes are assumed to involve the holding of goals and intermediate results in the limited-capacity working memory (see Chapter 9).

Gilhooly *et al.* (1999) studied individuals thinking aloud while solving the Tower of London problem. Their results suggest that working memory limitations tend to shape search patterns so that typically one action is selected from those available at each step. Search builds up a limited length of sequence before returning to the start state and re-exploring. (This process of searching depth first to a certain limit, then backing up and systematically searching all branches of the search tree to the depth limit is known as ‘progressive deepening’.) Gilhooly *et al.* found that the general strategy used was means–ends analysis, which generated a search pattern focused on reducing differences between the current state and the goal state. Means–ends analysis has been generally found to be the typical approach in the related Tower of Hanoi problem (Luger, 1976).

Similar results, indicating very focused mental search, have also arisen from studies of the hobbits and orcs task (Thomas, 1974; Simon and Reed, 1976) and the water jars task (Atwood and Polson, 1976). The state–space diagram for the hobbits and orcs task (Figure 10.5) appears below in Box 10.2 (overleaf).

Means–ends analysis typically involves reducing differences between the current state and the goal state, and so moves that bring the solver closer to the goal tend to be preferred. Thomas (1974) found that participants solving the hobbits and orcs problem found the transition from State 110 to State 221 (see Figure 10.5) especially problematic. The move involves bringing back one hobbit and one orc, which seems at odds with the general strategy of moving closer and closer towards the goal.

The water jars problem has also been used to examine search in problem solving. The problem requires participants to find a way of moving water between jars of given capacities from a starting state in which the largest is full to a goal state in which the water is distributed in a particular way over the three jars.

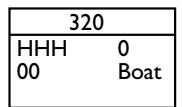
Response data from both the hobbits and orcs and the water jugs problems suggest a model in which solvers look ahead, evaluating a few possible steps at each point in terms of whether they appear to lead to new states closer to the goal or not. For example, at State 331, in Figure 10.5, a solver could look ahead to the next move and evaluate two possibilities, which would move them either to State 220 or State 310. Solvers appear to prefer moves that seem to take them closer to the goal, repeating the entire procedure until the goal is reached (Jeffries *et al.*, 1977). The preference for new states is a heuristic to avoid looping (or revisiting old states in a particular sequence, or cycle), as it is all too easy in these tasks to go round in circles! Avoiding loops is also important in the Tower of London task and Davies (2000) found that participants did not simply rely on

## 10.2

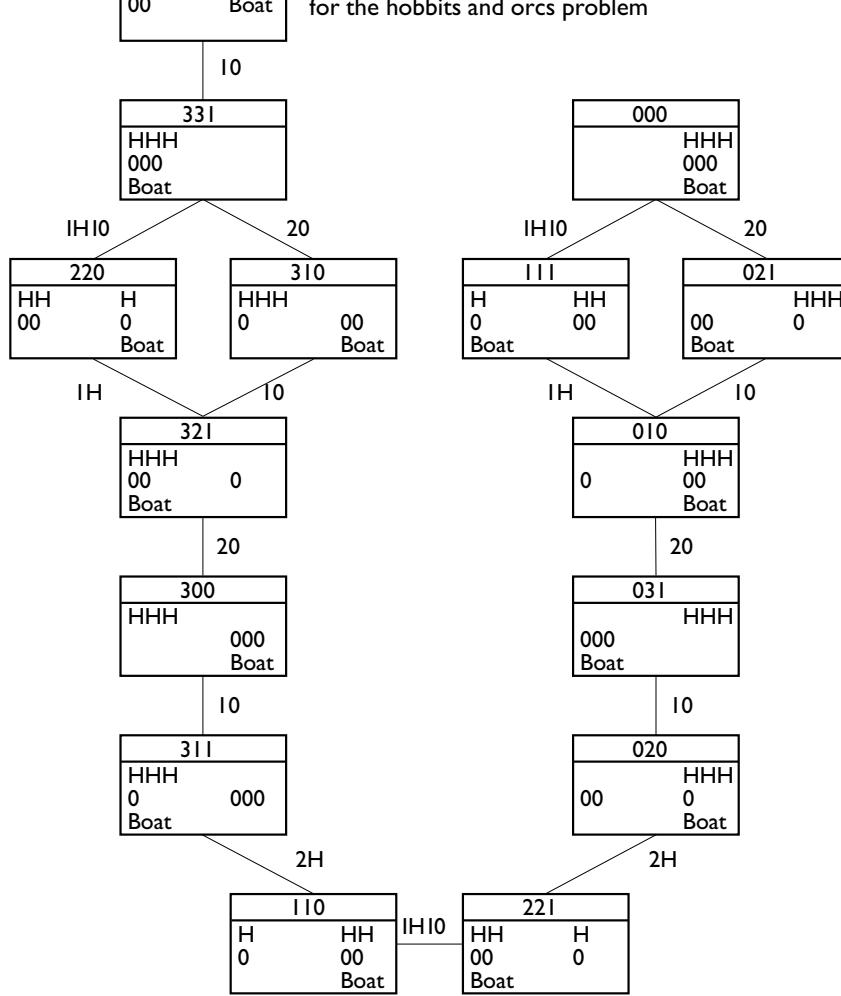
### The hobbits and orcs task

The hobbits and orcs task requires participants to find a way of transporting three hobbits and three orcs safely across a river in a boat. The boat can only hold two creatures at a time and on either side of the river the orcs must never outnumber the hobbits at any time.

In this state-space diagram (see Figure 10.5), each box represents a single state of the task. The number of hobbits (H) and orcs (O) on the left- and right-hand side of each box indicate the number of hobbits and orcs on the left and right banks of the river at any given time. Each state is labelled with a three-digit number, with the first digit representing the number of hobbits, the second the number of orcs and the third the number of boats all on the left bank of the river. So, for example, State 331 (near top left) indicates the start of the problem, with three hobbits, three orcs and the boat all on the left bank of the river. State 000 (top right) is the solution state, with all six creatures and the boat transported to the opposite side.



**Figure 10.5** State-space diagram for the hobbits and orcs problem



memory to judge whether a state had been encountered previously but would seek to infer whether it could have been a precursor of the present state. If the state in question could not have been a precursor then it was safe to assume that it was new.

## 2.4 Information processing approaches to insight

Sometimes, a change of representation may be required in order to solve problems that initially induce unhelpful ways of representing the problem. As we discussed earlier, the Gestalt school in the 1920s and 1930s regarded such re-structuring as the basis of insight into difficult problems. Though the Gestalt approach was criticized for a lack of clarity in explaining how re-structuring took place, Ohlsson (1992) offered some suggestions as to how restructuring might occur.

Ohlsson (1992) proposed that when working on a problem people generate possible actions or **operators** from long-term memory, which are cued by the problem representation. Applying the operators to the current problem state leads to a new problem state, which in turn elicits further possible operators. In this way, a problem with a useful initial representation may be solved as eventually a state is reached from which cued operators lead to the goal. However, if the initial representation is misleading, then a state will be reached in which no new useful operators can be retrieved. Subjectively, we experience this state, labelled an **impasse**, as a mental ‘blank’; we cannot think of anything new to try. Ohlsson suggests that impasses can be overcome by changing the problem representation so that new operators can be cued, retrieved and tried out. Specifically, Ohlsson proposed three ways in which the problem representation could be changed or re-structured and these are (1) **elaboration**, (2) **re-encoding** and (3) **constraint relaxation**.

Elaboration involves adding information to the representation by observing previously unnoticed features. For example, to use the matchbox tray in solving the ‘candle’ problem (Duncker, 1945), discussed earlier, the solver has to notice the possible use of the tray as a platform.

Re-encoding involves changing the encoding rather than simply adding new information. For example, how could a man have legally married 20 women in one month in a country where polygamy is illegal and none of the women have died or been divorced? To solve this, we have to change the encoding of ‘married’ from the man becoming husband to each woman to the man causing others to become married to each other, for example, the man in question could be a minister who is entitled to perform marriage ceremonies.

Constraint relaxation involves making the goal requirements less restrictive than initially assumed. For example, one source of difficulty in the 9-dot problem is the tendency to over-restrict the goal so that the four lines are kept within the square array of dots; removing (or relaxing) this constraint is necessary for a solution. Chronicle *et al.* (2001) found that another major source of difficulty with the 9-dot problem is that people tend to apply a heuristic search process for lines that cancel as many dots as possible; such a search process generates three lines within the square and the failure of the approach would only be realized with an extended look-ahead. Chronicle *et al.* argue that Ohlsson’s re-encoding

(where the problem representation is altered) is required: hence, the general failure of attempts to facilitate performance by giving hints about the need to move outside the square.

A further demonstration of the role of constraint relaxation comes from studies by Knoblich *et al.* (1999) of matchstick algebra problems. Examples of these types of problem are as follows: make the equations below, involving Roman numerals, true by moving a single stick (see Figure 10.6).

Type A:  $\text{V} \mid = \text{V} \mid \mid + \mid$

Type B:  $\mid \text{V} = \mid \mid \mid - \mid$

**Figure 10.6** Matchstick algebra problem

There are large differences in solution times and rates for these types of problems, with Type A being easier than Type B. In the Type A example, the rightmost stick from the ‘ $\text{V} \mid \mid$ ’ is moved to the right of ‘ $\text{V} \mid$ ’. This gives ‘ $\text{V} \mid \mid = \text{V} \mid + \mid$ ’, which is true. In the Type B problem, the operators ‘=’ and ‘-’ are changed by moving one of the horizontal sticks from the ‘=’ and placing it over the ‘-’ to make the true equation ‘ $\mid \text{V} - \mid \mid \mid = \mid$ ’. Knoblich *et al.* argue that it is harder to break the constraint on changing operators than on changing number values.

## Summary of Section 2

---

- Problem solving begins with a problem representation.
- The information processing approach analyses problem solving in terms of search within the space of possibilities arising from a particular way of representing the problem.
- Manipulating instructions and the appearance of problems appears to influence the nature of the problem representation constructed, which in turn can affect the ease with which the problem may be solved.
- Some representations of supposedly ‘simple’ problems render the problems extremely difficult to solve.

### 3 Analogical problem solving

As we saw in the previous section, some problems we encounter are novel and difficult to solve even when they require a minimum of background knowledge and experience. However, we often encounter problems that are rather similar to problems we have tackled previously. Even if a solution to a new problem is not known, we may know and be reminded of the solution to similar problems and be able to use that known solution to suggest a solution to the new problem. That is, we may be guided to solution by the use of an analogy.

#### 3.1 Analogies in problem solving

Spellman and Holyoak (1992) found that experimental participants readily accepted analogies of the kind often used in discussions of international politics. For example, when Saddam Hussein attacked Kuwait in 1990, many commentators likened him to Adolf Hitler and George Bush Snr to Winston Churchill. Similarly, some regarded the ‘Domino’ theory – the US government’s belief that if one Asian country fell to communism then others would quickly follow – as compelling justification for the war on Vietnam.

In science, analogies have often been used to develop understanding. For example, the heart has been seen as analogous to a water pump and atomic structure as similar to that of the solar system. In cognitive psychology, we hope that analogies between human and computer information processing will likewise prove useful.

Studies of analogy use in problem solving have often used Duncker’s X-ray problem, which you met in Section 2, as the target problem. You may remember that the problem was to use rays to destroy a tumour in the centre of a body without destroying healthy tissue. The solution was to converge a number of weak rays on the tumour, which would then have a cumulative effect at that point. Gick and Holyoak (1980) gave their participants an analogous (or ‘base’) story about a general seeking to take a castle who had to divide his forces into small groups who then attacked the castle simultaneously on all sides. Participants were then given the X-ray problem, with and without a hint. The hint indicated that solvers should try to use the story to solve the X-ray problem. A control group did not receive the analogue before tackling the X-ray problem. Rate of solving was low for participants who had not been given the analogue, somewhat higher when the analogue alone had been given and markedly higher for the analogue followed by a hint.

Later studies (Holyoak and Koh, 1987; Keane, 1988) indicated that the closer the base story is in surface features to the target problem, the more transfer is likely. For example, Keane found that a very close analogy of a surgeon treating a brain tumour by radiation was much more often retrieved and used in tackling the X-ray problem, even after a week’s delay, than was the more remotely analogous story of the general dividing his forces.

Anolli *et al.* (2001) found that retrieval of a remote analogy was ineffective in itself without provision of a hint that the analogy contained a useful clue to solution. In a series of seven studies little benefit in terms of solving was found

for simply reminding participants of the analogous story without a hint to use the story. However, retrieval plus a hint was very effective.

Dunbar (2001) has noted that there is an *analogical paradox*; the paradox is that in real life use of abstract analogies that depend on deep structural similarities is common, while in laboratory studies participants tend to use superficial features and have difficulty with deeper forms of analogy. Blanchette and Dunbar (2000) found that participants who were asked to produce analogies that could be used in arguments about whether or not drastic cuts should be made in public spending during a budgetary crisis readily produced deep analogies that drew on a range of content areas having little superficial resemblance to political matters. Blanchette and Dunbar proposed that generating analogies requires participants to use structural rather than superficial features; also, the subject matter of the naturalistic analogy studies has been familiar and understood in some depth. In typical laboratory studies, the material is not highly familiar and there is little pressure to encode the base story in a deep way. A possible interpretation of Anolli *et al.*'s (2001) results (namely, target problem solution is increased by reminding of the base story plus a hint that the analogy contained a useful clue to solution) is that the hint encouraged a deeper structural representation of the base story, which in turn facilitates application of the analogy.

### 3.2 How do analogies work?

The first detailed theory of how people apply analogies in problem solving was the ‘structure-mapping’ theory (Gentner 1983; Gentner and Markman, 1997; Gentner *et al.*, 2001). According to this theory, there is a process of **analogical mapping** whereby a **structural alignment** is established between the representations of the base and the target. That is to say, explicit correspondences are established between the represented elements and relationships in the two situations. As an example, consider the solar system analogy of the atom. Typically, the solar system would be represented as having two types of object (the sun and the planets) and these exhibit various properties and relationships (e.g. sun is more massive than planets; planets orbit the sun). The analogy would align the sun with the nucleus of the atom and the planets with electrons. Aspects of the solar system model that do not map are omitted (e.g. no equivalent of moons in the atom). Higher order relationships also guide the alignment process. Thus, there is a higher order relationship such that less massive objects orbit more massive objects; this guides the inference from the solar system analogy that the electrons revolve around the nucleus. Falkenhainer *et al.* (1986) have successfully implemented the detailed model as a computer model known as the ‘structure mapping engine’ (SME).

Gentner and Gentner (1983) showed that different base analogies were common for understanding electrical flow. These influenced how well people solved different problems regarding electrical flow through circuits with batteries and resistors arranged in parallel or in series. The main analogies were electricity as a flow of fluid through pipes or as a flow of crowds through passageways. Example differences between the two analogical mappings are that in the fluid analogy electrical resistance would be mapped to pipe width while in the

crowd analogy resistance is mapped to gates in the passageway. People who used the fluid analogy performed better on battery problems; people who used crowd analogies were better on resistor problems. People pre-trained on both the analogies also showed similar results. A number of models similar to the structure-mapping theory have been proposed including the ‘analogical constraint mapping engine’ (ACME) (Holyoak and Thagard, 1989) and the ‘incremental analogy machine’ (IAM) (Keane, 1994).

## Summary of Section 3

---

- Research on analogical problem solving illustrates a paradox: we often cannot help but be reminded of problems similar to one we presently face. However, we often fail to see the crucial relationships between a current problem and one we have previously encountered.
- Representation seems to be at the core of the paradox.

## 4 ‘Complex’ problem solving

While research on the ways in which we solve puzzles and analogies has mapped out the terrain to a certain extent, it should be apparent that understanding how people solve the relatively ‘knowledge-lean’ problems we have looked at so far is only a part of the picture. Many problems require a considerable amount of knowledge if they are to be solved successfully. This section focuses on ‘complex’ problem solving, or problem solving that requires an extensive knowledge base. Does knowledge of a domain affect problem representation, and hence the likelihood with which a problem will be solved? Do problem representations change as knowledge is acquired and as skill develops? Can we characterize the development of skill in problem solving? Researchers have turned their attention to how experts and novices solve problems in their attempts to try to answer these questions.

While it may seem obvious that experts know more than novices, until relatively recently the layperson’s view of the expert might well have encapsulated the view that experts owe their skill to superior mental capacities, rather than to a vast body of specialist knowledge. The shift in emphasis began with some ground-breaking research on chess skill. We shall examine this work in some detail, because much of the later research on expert and novice problem solving developed from the early chess studies, and because findings obtained in later studies tend to echo those of the chess experiments described below.

First, some words on terminology. Researchers have examined both **adversary** and **non-adversary** problem solving. Chess play is an example of adversarial problem solving, because the game of chess involves an opponent. Code-breaking, de-bugging computer programs and medical diagnosis are examples of non-adversarial problem domains. Those engaged in adversary problem solving then must consider not only their own possible actions, but also those of an opponent.

## 4.1 The role of knowledge in expert problem solving

### 4.1.1 Early chess studies

De Groot (1946/1965) carried out a series of now classic studies of chess players. These were extremely significant and heralded the start of a new emphasis on knowledge in skilled problem solving. Information processing had taken centre stage as the dominant paradigm and many researchers busied themselves with the construction of models of cognitive processors and processes. Up until then, it had been implicitly, if not explicitly, assumed that skilled problem solvers must have superior information processing capabilities. De Groot tested this assumption in a novel way.

De Groot asked five grand masters (the highest skill level attainable in chess) and five skilled players to think aloud as they studied a chessboard and chose a move. If information processing capacities are a key determinant of expertise, then we would expect to find the grand master players, with their superior capacities, searching further ahead and conducting broader searches for candidate moves. The evidence from the think-aloud protocols however was surprising and revealed no reliable quantitative differences at all between the grand masters and the highly skilled players. The only difference that did emerge between the two groups was unremarkable – the grand masters ultimately chose better moves.

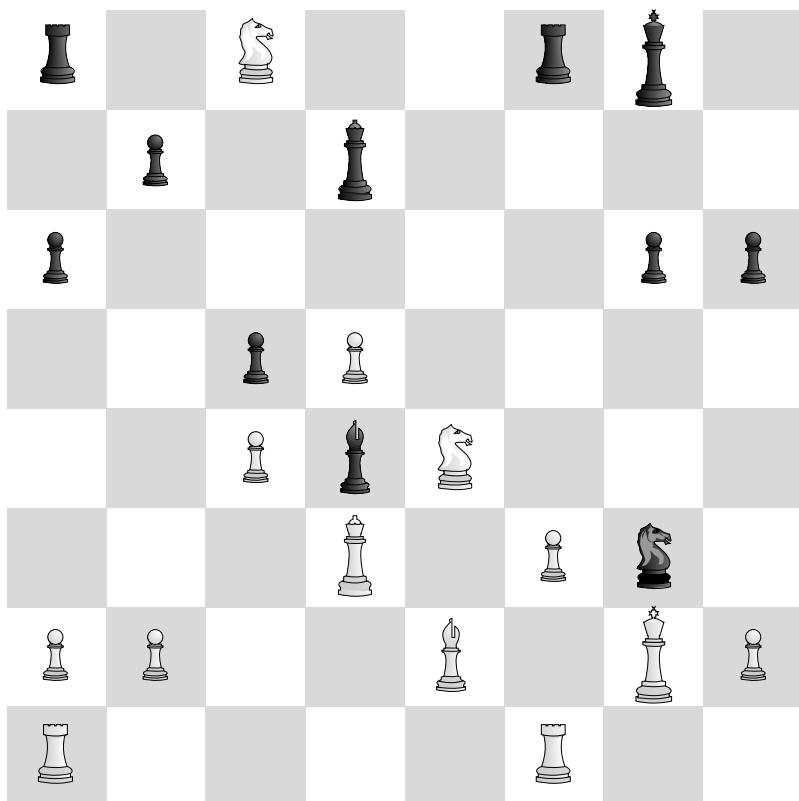
De Groot also employed what is known as a ‘recall-reconstruction’ paradigm (see Box 10.3). He showed chess players chessboards with pieces arranged from actual games. The boards were presented to players for 2–15 seconds, and then removed. He then asked the chess players to reconstruct the board positions from memory. The chess masters could reconstruct the boards almost without error (91 per cent of pieces correctly replaced), whereas the poorer players averaged only 41 per cent correct. Skill level then was linked to the amount of information remembered about the chessboard positions.

Chase and Simon (1973b) devised a second task, where players had to reconstruct a chessboard while the board they had to correctly match was still in view. Although this may seem an odd task, the point was to find out how many pieces were placed on the target board after each glance, what those pieces were, and how much time elapsed between placing pieces on the board. Chase and Simon found that the strongest chess players replaced more chess pieces on the board following each glance, replaced pieces more quickly and tended to replace pieces together that bore some meaningful relationship to each other than did the weaker players. These findings suggested that experts not only possess more knowledge about their domain of expertise, but that their knowledge is organized in more meaningful and readily accessible ways.

These early studies of chess skill showed that skill depended at least in part on the acquisition of domain knowledge, and stimulated a vast amount of research on the nature of expert problem solving and the relationship between knowledge and skill. We summarize some of the key studies below. These studies sought to characterize the empirical phenomena associated with skill in problem solving, phenomena that theories of skill acquisition and problem solving would ultimately have to accommodate and explain.

**10.3****Research study****The recall-reconstruction paradigm in chess**

Chase and Simon (1973a) extended the basic chessboard recall-reconstruction paradigm, originally used by Lemmens and Jongman in an unpublished study of 1964. In one study, Chase and Simon presented boards with between 20 and 22 chess pieces arranged on them to three chess players (a master, a class A [highly skilled] player and a novice). Some of the boards were presented with chess pieces arranged as they might be in a real game (see Figure 10.7), while others were presented with the chess pieces arranged at random.



**Figure 10.7** A chess board position from a real game

Source: adapted from Chase and Simon, 1973a

Players were given five seconds to study each board. The board was then covered up, and each player was asked to reconstruct on another board the position just seen. Results from the first memory trial showed that the master player was much better (16 pieces replaced correctly) at accurately replacing the chess pieces than both the Class A player (8 pieces replaced correctly) and the novice player (only 4 pieces replaced correctly). The skilled players' advantage only held for chessboards with pieces placed in plausible, real-game positions. When the different players were asked to reconstruct the random boards from memory, all players performed equally poorly, correctly replacing only a small proportion of pieces. This suggested a connection between memory for meaningful patterns and problem-solving skill.

#### 4.1.2 Experts work forwards

Larkin *et al.* (1980) were interested in possible strategic differences between experts and novices. Experts know more than novices, but do they also use qualitatively different problem-solving strategies to novices? They asked expert and novice physicists to solve a range of physics problems. Using protocol analysis, they found that experts tended to use a **working forwards** strategy, beginning with information given in the problem statement and using that to derive a solution. Novices, on the other hand, used a **working backwards** strategy (means–ends analysis), starting with the goal, or quantity to be solved, and working backwards from that to the given information, until they were able to solve one part of the problem. Novices then typically re-traced their steps, working forwards until the problem was solved.

Why do experts and novices use different problem-solving strategies? It appears that experts use their domain knowledge to generate a good problem representation, which supports the use of a working forwards strategy. In the absence of detailed knowledge about the relationships between variables relevant to the problem, novices seem to have no option but to fall back on means–ends analysis, or even trial and error learning.

#### 4.1.3 Experts have better problem representations

Chi *et al.* (1982) tackled the issue of problem representation and categorization by experts and novices. Experts know more and they use distinct problem solving strategies. Is expert problem solving also supported by more effective ways of representing and categorizing problems? Chi *et al.* asked expert and novice physicists to think aloud as they categorized physics problems on the basis of similarities in terms of how the problems might be solved. Unusually then, participants did not actually have to solve the problems.

The two skill groups did not differ on quantitative measures, such as number of categories or time to categorize. This showed that novices were not limited in their capacity to discriminate problems. However, there were clear qualitative differences in the nature of the categories into which problems were sorted. Novices referred to objects and key words contained in the problem (such as ‘levers’ and ‘pulleys’), and appeared to use these irrelevant ‘surface structure’ details as a basis for categorization. Experts, on the other hand, referred to the physics principles and laws (the ‘deep structure’) that were needed to solve the problems in their justifications. Problems that could be solved by reference to the same principle or law were perceived by the experts to be similar and were grouped together. Novices tended to group together problems that were similar in ‘surface structure’, while experts sorted problems on the basis of similarity in ‘deep structure’. It seems then that experts are aware of commonalities between problems in terms of how they might be solved.

Schoenfeld and Herrmann (1982) carried out a rather similar study, looking at mathematical problem categorization among mathematics professors and novices. Their participants read through the set of problems and then grouped together those problems they considered to be mathematically similar. The study confirmed the findings of Chi *et al.* (1982), with novices sorting the problems on the basis of superficial details, or surface structure, and the professors sorting problems on the

basis of similarities in solution methods, or deep structure. (You will notice some similarities here with the discussion in Chapter 5 of how different groups of people categorized trees.)

Chi *et al.* found in their study that experts were able to perceive an appropriate solution method within 45 seconds. This suggests that knowledge useful for a particular problem is accessed, or becomes available, when a problem is categorized as a specific type. These categories may correspond to problem schemata or ‘packets’ of knowledge that can be used to solve a particular type of problem.

#### **4.1.4 Experts become expert through extensive practice**

It is often said that ‘practice makes perfect’. In the context of problem solving, researchers noticed many years ago that performance improves with practice in a very systematic and predictable fashion. What is particularly interesting is the observation that, regardless of what is being learned, performance improves with practice in a highly predictable way. The relationship, known as the ‘power law of practice’, has been known for a long time, though there is an ongoing debate as to whether practice learning data are best fit by a power function, or some other function. The relationship shows up in Snoddy’s (1926) study of mirror-tracing of visual mazes. It appears in perceptual tasks such as Kolers’ (1975) studies on mirror-reading (where text is transformed), in pattern recognition (Neisser *et al.*, 1963), and in tasks from the domain of human computer interaction (e.g. Card *et al.*, 1983).

Practice then seems to be a factor in the development of skill. The improvement in performance with practice applies over a wide range of activities that are better described as ‘tasks’ and which include problem-solving tasks, as well as other kinds of tasks (for example, juggling and search tasks). Why does performance improve with practice? Three main classes of explanation have been proposed:

- 1 Individual task components are executed more efficiently.
- 2 Sequences of task components are executed more efficiently.
- 3 Qualitative changes occur in representations of task structure.

The first two explanations argue that performance improves with practice because the piece-meal recovery of declarative knowledge into working memory is reduced, and because we learn to run off sequences of procedures in ever greater units or chunks. The third explanation asserts that performance improves because the nature of the task changes, either because the task is restructured or because we shift from algorithm-based to memory-based processing (an example of the latter is Logan’s [1988] ‘instance’ theory of automaticity).

How much practice is needed to achieve excellence? Ericsson *et al.* (1993) have given ten years as a ballpark figure for attaining high levels of performance in a variety of areas (e.g. chess, mathematics and violin playing). In a review of the literature on practice and performance, Ericsson (1991) has suggested that it takes at least ten years to reach the international level of performance in sports, the arts and sciences. Simon and Chase (1973) estimated that it took some 3,000 hours practice to become an expert and around 30,000 hours to become a chess master. The preparation period may often commence at an early age, possibly because it takes so long to acquire the necessary knowledge. While it clearly takes a long time to attain

very high levels of performance, it is nevertheless possible to train subjects to improve on their previous best performance. Ericsson and Harris (1990) trained an individual who was not a chess player over a period of 50 hours to recognize chess positions almost as accurately as some chess masters.

However, as Ericsson and Polson (1988) found, practice itself is not a guarantee of superior performance. In their study, the waiter most skilled in remembering orders used more effective encoding strategies and achieved much better performance than his equally experienced counterparts, who did not use the same optimal encoding strategies to remember dinner orders. This means that something else must mediate between practice and performance.

What appears to be critically important is not how much practice individuals have, but what they actually do while they are practising the skill. If it takes a very long time to become expert, then clearly we need to document what individuals do over a longer time scale than is usually considered. We return to this point in Section 5.2 when we explore individual differences in problem-solving performance.

## 4.2 A modal model of expertise?

The early chess studies triggered a vast amount of research that used what became known as the ‘expert–novice’ paradigm. The model of chess expertise that emerged became known as the ‘pattern recognition hypothesis’, because it assumed that skilled performance depended upon the ability to access previously learned patterns, such as configurations of chess pieces on a board, from long-term memory.

The general idea that performance depends upon a large body of highly structured domain-relevant knowledge and skill has been borrowed by researchers examining skill-related differences in non-adversarial domains such as physics (Chi *et al.*, 1982), mathematical problem solving (Schoenfeld and Herrmann, 1982), computer programming (McKeithen *et al.*, 1981) and political science problem solving (Voss *et al.*, 1983). Results consistently showed a link between expertise and knowledge, suggesting that a ‘modal model’ of expertise was emerging, whereby expertise depends upon the acquisition and organization in long-term memory of domain-relevant knowledge and skill.

Although supported by the data, these initial observations about expertise seem descriptive and lack explanatory power. Sternberg (1995) is but one researcher to have commented upon this. Over-use of the paradigm appeared to constrain the nature of the findings to a series of observations about experts ‘knowing more’ than novices. Accounting for these findings was nonetheless a challenge for theories of skill acquisition, but many researchers recognized that there was more to expertise than the gradual accumulation of domain-specific knowledge.

As we shall see in the next section, when researchers began to explore different questions about the nature of expertise and about skill acquisition, some findings emerged that challenged the prevailing view of expertise while at the same time yielding valuable insights into the relationship between memory and skill, and the development of expertise itself.

## Summary of Section 4

---

- On memory tests for information from their domain of expertise, experts remember more than novices.
- Experts are superior to novices in knowledge rather than in basic capacities.
- Experts use a working forwards strategy, while novices tend to work backwards.
- Experts construct better problem representations than novices.
- Experts become expert through extensive practice.

## 5 Prospects for problem-solving research

In this section, we focus on research that points to some limitations of the general model of expertise outlined above, and we go on to discuss some of the directions research in problem solving has been taking.

### 5.1 Does expertise transfer?

It is perhaps ironic that early indications that all was not well with the modal model of expertise came from research on chess skill, which had originally played such a large part in stimulating research on expert and novice problem solving.

#### 5.1.1 Chess skill and memory

The classic chess recall-reconstruction experiments (as discussed in Box 10.3 above) showed that the master chess players' memory advantage held only for meaningful chess positions, suggesting that memory determines chess skill. However, subsequent studies of expert chess play question this conclusion and suggest that memory cannot be the sole determinant of skill (Holding, 1985).

For instance, Holding and Reynolds (1982) sought to determine whether skill differences could be shown in the absence of differences in memory. They asked players differing in their skill ratings to memorize random positions. Next, players were asked to select the best continuation moves. Skill level was unrelated to recall of random positions, replicating the findings of de Groot (1965). However, the interesting finding is that the number of best moves chosen correlated positively with playing strength. Therefore, differences in memory for chess patterns cannot account for the finding that better players chose more good moves from random starting states. This suggests that, for highly skilled chess players at least, something other than memory for highly familiar configurations of chess pieces may be implicated in chess skill. An additional factor is likely to be the ability to evaluate a given position.

Holding (1979) set out to examine the relationship between skill level and evaluation among chess players. He presented fifty players varying in skill level (from Class A players, the strongest, to Class E players, the weakest) with a set of test positions and asked them to indicate which side had the advantage, and to rate the strength of the advantage. The results confirmed that the ability to evaluate chess

positions is an important dimension of chess skill. Stronger players were more often right about the outcomes of the games from which the test positions were taken. Also, the subjects were asked to suggest what they thought was the best move in each position. The average number of times that the players' move choices corresponded with the grand master move in the actual games varied systematically with rating class (A: 3.6; B: 3.0; C: 2.9; D: 2.3; and E: 1.6). Therefore, the higher rated players made more good moves and fewer evaluation errors.

### 5.1.2 The role of general and specific methods

Schraagen (1993) carried out a more detailed examination of the problem-solving performance of different groups of experts and novices. Most studies of expertise have shown that experts draw upon a large body of domain knowledge when asked to solve a problem from their domain of expertise. Anderson's (1983) ACT\* theory predicts that when domain knowledge is lacking, experts should fall back on general strategies, or 'weak methods'. Schraagen asked his participants to design an experiment in the area of sensory psychology. He compared the reasoning of domain experts (psychologists with around 10 years' experience in designing experiments in the area of sensory psychology) with 'design' experts (psychologists with around 10 years experience in designing psychology experiments in general). The problem facing subjects was to design an experiment to investigate what people taste when they drink a given brand of cola. While domain knowledge was important (the domain experts generated better solutions), the form of the design experts' reasoning was comparable to that of the domain experts. When knowledge is lacking, it seems that there may be skills of intermediate generality that do transfer. These findings are at odds with theoretical frameworks that argue for the domain-specificity of expertise.

Schunn and Anderson (1999) carried out a similar study to examine whether expert scientists from different domains shared some skills. They asked domain-experts (psychologists skilled in designing memory experiments), task experts (psychologists skilled in areas other than memory research), and undergraduate students studying different courses to think aloud while designing an experiment to investigate an unexplained aspect of memory. Analysis of protocols and performance data showed that the domain-experts designed the best experiments. Domain-experts and task-experts differed in terms of domain-specific skills, while task-experts and undergraduates differed on domain-general skills. Through the analysis of verbal protocols, the researchers were able to identify a much larger set of domain-general skills that are important in scientific reasoning.

## 5.2 Individual differences

The expert–novice paradigm contrasts the performance of experts and novices solving the same set of problems drawn from a given domain. Although the problems used are likely to be non-trivial to the novices, they scarcely present a problem in any meaningful sense to the experts. This is necessarily so because if the problems were truly challenging for the experts, novices would not be able to even begin to solve them. However, if the experts have not really been taxed with a 'problem', have we learned anything at all about expert problem solving? The expert–novice paradigm also tends to imply that novices know nothing, or know

little of relevance. As we shall see, novices do not approach novel problems with ‘empty heads’. They bring to bear whatever knowledge and strategies they are able to and, in doing so, it is clear that some novices are better learners than others. We shall now examine the extent to which implicit assumptions about the homogeneity of both novice and experts groups are reasonable.

### 5.2.1 Are all learners the same?

Novices have tended to be described in terms of what they do not have, or do not do. A more positive approach is to examine what novices *can* do, and ways in which they differ. In so doing, this work shifts the emphasis from problem-solving performance to learning and the acquisition of skill in problem solving. Models and theories of problem-solving performance must not only account for differences between skilled and less-skilled individuals, they must also explain how skill is acquired.

Some interesting work has examined differences between good and poor learners, and this has shed some light on what might mediate between practice and performance. Many of us will have noticed that people tend to differ in rate of learning. While it is an over-simplification to suggest that novices start with a blank slate, most novices begin from a position of not having much of the skill in question. If knowledge relevant to the skill does not mediate or support their performance early on, then what does?

Thorndyke and Stasz (1980) examined learning strategies differentiating good from poor learners of map information. Good learners used more efficient techniques for encoding spatial information, more accurately determined what they knew and what they had yet to learn, and were better able to focus their attention on map elements they had not yet learned. Green and Gilhooly (1990) conducted a similar experiment, studying novices learning to use a statistical package on a mainframe computer. Good learners tended to adopt an exploratory approach to learning, made better use of worked examples from handouts and evaluated their learning. Slower learners tended to over-use worked examples, generated and tested more erroneous hypotheses and seemed to either ignore, or fail to use, error feedback. Both these studies suggest that good learners make effective use of **metacognitive** processes and strategies.

Chi *et al.* (1989) and Chi *et al.* (1994) have been especially interested in the role played by explanation in learning and, in particular, whether novices may be distinguished by the extent to which they generate explanations while solving problems. In their studies, they equated students for background knowledge (of physics and biology) and then analysed the think-aloud protocols students produced as they studied the problems. In one study, good learners seemed to spontaneously self-explain more than poor learners. Good learners used the examples they had studied to check their solutions whereas poor learners used the examples to help them to find solutions. Chi *et al.* (1994) showed that prompting students to self-explain as they studied led to better problem solving than simply asking students to study the materials. Renkl (1997) showed that the self-explaining effect is not simply due to some students spending longer studying. In a study that controlled for time-on-task, Renkl found that quality of self-explanations reliably predicted learning success. Generating self-explanations then, whether spontaneously or in response to

a prompt to do so, seems to serve an elaborating role in early learning, aiding understanding and schema development. Schema development is of course central to skill development.

### 5.2.2 Can we enhance the rate of skill acquisition?

Sweller and his colleagues (Sweller, 1988; Sweller *et al.*, 1983) have demonstrated that schema acquisition can be retarded by the use of means–ends analysis. Paradoxically, the very strategy that novices appear to rely on in early learning (recall Section 4.1.2, and the study by Larkin *et al.*, 1980) has been shown to inhibit knowledge acquisition. Sweller *et al.*, hypothesized that an emphasis on a goal (which occurs with the means–ends analysis strategy) might overload the system, leaving few resources available for inducing relevant schematic knowledge. They tested this hypothesis by de-emphasizing the goal in a set of kinematics (a branch of physics) problems given to one group of novices. For example, one problem ended in the following way: ‘In 18 sec a racing car can start from rest and travel 305.1 m. Calculate the value of as many variables as you can’.

A second group of novices received the same problems, but the final sentence was altered to include a specific goal:

‘In 18 sec a racing car can start from rest and travel 305.1 m. What speed will it reach?’

Participants who were given the no-goal problems switched more swiftly to a working forwards strategy than did novices who were given the goal problems. One interpretation of these findings is that the presence of a goal biases individuals towards the use of a means–ends strategy, which imposes high processing demands. This would have the effect of reducing the available resources for acquiring knowledge about the relationships among principles. De-emphasizing the goal then could work by reducing working-memory load, thereby freeing resources. This would facilitate schema acquisition, thereby enhancing learning.

There is an alternative explanation for the facilitating effect of reduced goal specificity on learning. Vollmeyer *et al.* (1996) examined the effects of goal specificity and systematicity of learning strategy in learning and transfer within a complex dynamic system. Their findings were consistent with Sweller’s claim that general problem-solving methods might enable a person to attain a specific goal, but do not promote learning of the overall structure of a problem space. Burns and Vollmeyer (2002) have taken this work further, and have shown that non-specific goals seem to aid learning by encouraging more hypothesis testing. Their work shows that it seems to be hypothesis testing, rather than the reduction in goal-specificity, that encourages learning.

There may be another dimension to the effects of goal specificity on learning and problem solving. Green (2002) argued that reducing goal specificity also has the effect of altering the way in which a problem comes to be represented. In her experiment, it was the nature of the problem representation that was crucial to performance, rather than the reduction in goal specificity. Green also points out that it is important to distinguish learning from problem solving. Instructions that led to swift learning seemed to result in poor problem solving, while instructions that seemed to lead to slower learning paid off later on by giving rise to better problem

solving. Different instructions influence the nature of the task or problem representation, and this in turn affects both learning and problem solving performance. This echoes the point we made earlier with regard to the impact of internal and external representations upon learning and problem solving.

Some recent studies by Haider and Frensch (1996, 1999a and b) have focused on ways in which we learn to ignore task-irrelevant information, and process only task-relevant information. Haider and Frensch have shown that as we become more skilled, we typically learn to ignore redundant information. Not all individuals behave in the same way though. In one of their studies, they found that some individuals fail to reduce the amount of information heeded, even after extended practice. Green and Wright (2003) have extended these findings, examining what happens when two information sources associated with one event are presented. When individuals have a choice of information sources relevant to the task, they tend to prefer to use the first encountered source. Information reduction then serves to reduce processing of task-irrelevant, as well as duplicated (but possibly task-relevant) information. The assumption that we come to process less information is at odds with some theories of skill acquisition like ACT\* and ACT-R (which you will explore in more detail in Chapter 16).

The studies we have discussed in this section provide some clues as to how individuals learn to solve problems more effectively. What is apparent is that learners do not all behave in the same way. Certain learning procedures and strategies facilitate knowledge acquisition, and there is evidence that problem representation again plays a key role.

Novices seem to differ from each other then, and their rates of learning can vary. Do experts form a homogeneous group? We examine this question now.

### 5.2.3 Do experts differ?

It is sometimes tacitly assumed that experts form a homogeneous group, with considerable overlap among experts in what they know. If this is the case, then we may safely generalize from studies of experts, and talk about ‘typical’ expert problem solving behaviour. However, it is likely that the assumption of homogeneity is at best an over-simplification. Draper (1984) carried out a study of expertise in UNIX (a computer operating system that uses brief commands, often in the form of consonant strings, e.g. ‘LS’ [lists all the files in the current directory]). He found that UNIX users share some knowledge of UNIX commands, but mostly they use different commands from each other. Further, the size of their vocabulary of commands varies greatly from individual to individual. If it was simply the case that experts knew more than novices, then we might expect the expert user’s vocabulary of UNIX commands to subsume the novice user’s vocabulary. Draper’s results show that this is not the case. In fact, there is very little overlap between the novice and the expert users’ vocabularies. Novices do not all use the same commands as each other, and neither do experts. Draper argues that UNIX experts are better seen as specialists with a subset of UNIX commands.

What have we learned from such studies? Firstly, we see that ‘experts’ differ among themselves. This point was also made by Charness (1991), in his examination of chess skill, who found that chess masters do not know the full range of opening variations (there are some 50,000), middle-game combinations

(around 1,817) and end-games (some 8,500). Indeed, it is questionable whether they could actually learn the full set. Instead, chess masters specialize in a subset of each of the three classes. Secondly, we can see that there are different kinds of expertise. The physics expert, for example, is an expert in a domain where knowledge of the principles themselves is sufficient to solve most of the problems that may be encountered. The chess master and the UNIX expert exercise their skill in a domain where it is virtually impossible to learn all there is to know. There is not a body of ‘principles’ as such that are logically sufficient to solve most problems.

Chess, computer programming and physics are, nonetheless, well-defined domains. In the case of chess in particular, certain problem states can become highly familiar, and stronger players can capitalize upon their ability to recognize good problem states. Sometimes though, recognition hinders the construction of optimal representations.

## Summary of Section 5

---

- Neither experts nor novices form a homogeneous group.
- Skill in problem solving involves more than just the accumulation of knowledge.
- Problem-solving skill may be enhanced in a number of ways.

## 6 Conclusion

Nearly a century of research on problem solving has yielded some impressive findings. Important phenomena, such as insight and fixation, which have long taxed researchers are now amenable to more rigorous, systematic investigation thanks to methodological and theoretical advances, not to mention the advent of cognitive modelling. We now have a better understanding of analogical reasoning, helping us to appreciate how analogical reasoning occurs and why it sometimes fails. Advances have been made in understanding how we become skilled in solving problems from a wide range of complex domains, which have in turn led to a better understanding of expertise and learning.

We do not yet have a theory of problem solving; nor do we have a theory of learning, but progress is being made. Underpinning research on problem solving though is a recurring theme, and it is this: representation is fundamental to problem solving, just as it is fundamental to many other areas of psychology. Problem representation is likely to be influenced by many variables, some of which we have only begun to explore. We hope we have stimulated your interest in this area of psychology sufficiently that you will see the potential for problem solving research in terms of its wider application, as well as its theoretical significance.

## Further reading

- Chronicle, E.P., MacGregor, J.N. and Ormerod, T.C. (2004) 'What makes an insight problem? The roles of heuristics, goal conception and solution recoding in knowledge-lean problems', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.30, pp.14–27.
- Anderson, J.R. (2002) 'Spanning seven orders of magnitude: a challenge for cognitive modeling', *Cognitive Science*, vol.26, pp.85–112.
- Ericsson, K.A. and Kintsch, W. (1995) 'Long-term working memory', *Psychological Review*, vol.102, no.2, pp.211–45.

## References

- Adamson, R.E. and Taylor, D.W. (1954) 'Functional fixedness as related to elapsed time and set', *Journal of Experimental Psychology*, vol.47, pp.122–6.
- Anderson, J.R. (1983) *The Architecture of Cognition*, Cambridge, MA, Harvard.
- Anolli, L., Antonietti, A., Crisafulli, L. and Cantoia, M. (2001) 'Accessing source information in analogical problem-solving', *Quarterly Journal of Experimental Psychology*, vol.54A, pp.237–61.
- Atwood, M.E. and Polson, P.G. (1976) 'A process model for water jug problems', *Cognitive Psychology*, vol.8, pp.191–216.
- Blanchette, I. and Dunbar, K. (2000) 'How analogies are generated: The roles of structural and superficial similarity', *Memory and Cognition*, vol.28, pp.108–24.
- Boshuizen, H.P.A. and Schmidt, H.G. (1992) 'On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices', *Cognitive Science*, vol.16, pp.153–84.
- Burns, B.D. and Vollmeyer, R. (2002) 'Goal specificity effects on hypothesis testing in problem solving', *The Quarterly Journal of Experimental Psychology*, vol.55A, no.1, pp.241–61.
- Card, S.K., Moran, T.P. and Newell, A. (1983) *The Psychology of Human–Computer Interaction*, Hillsdale, NJ, Erlbaum.
- Charness, N. (1991) 'Expertise in chess: the balance between knowledge and search' in Ericsson, K.A. and Smith, J. (eds) *Towards a General Theory of Expertise: Prospects and Limits*, Cambridge, MA, Cambridge University Press.
- Chase, W.G. and Simon, H.A. (1973a) 'The mind's eye in chess' in Chase, W.G. (ed.) *Visual Information Processing*, New York, Academic Press.
- Chase, W.G. and Simon, H.A. (1973b) 'Perception in chess', *Cognitive Psychology*, vol.4, pp.55–81.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. and Glaser, R. (1989) 'Self-explanations: how students study and use examples in learning to solve problems', *Cognitive Science*, vol.13, pp.145–82.
- Chi, M.T.H., Glaser, R. and Rees, E. (1982) 'Expertise in problem solving' in Sternberg, R.J. (ed.) *Advances in the Psychology of Human Intelligence*, Hillsdale, NJ, Erlbaum.

- Chi, M.T.H., de Leeuw, N., Chiu, M-H. and LaVancher, C. (1994) 'Eliciting self-explanations improves understanding', *Cognitive Science*, vol.18, no.3, pp.439–77.
- Chronicle, E.P., Ormerod, T.C. and MacGregor, J.N. (2001) 'When insight just won't come: the failure of visual cues in the nine-dot problem', *Quarterly Journal of Experimental Psychology*, vol.54A, no.3, pp.903–19.
- De Groot, A.D. (1965) *Thought and Choice in Chess*, (original edition in Dutch, 1946), The Hague, Mouton.
- Davies, S.P. (2000) 'Move evaluation as a predictor and moderator of success in solutions to well-structured problems', *Quarterly Journal of Experimental Psychology*, vol.53A, no.4, pp.186–201.
- Draper, S.W. (1984) 'The nature of expertise in UNIX' in Diaper, D., Gilmore, D., Cockton, G. and Shackel, B. (eds) *Proceedings of Interact*, Amsterdam, Elsevier, pp.465–71.
- Dunbar, K. (2001) 'The analogical paradox: why analogy is so easy in naturalistic settings, yet so difficult in the psychology laboratory' in Gentner, D., Holyoak, K.J. and Kokinov, B. (eds) *Analogy: Perspectives from Cognitive Science*, Cambridge, MA, MIT Press.
- Duncker, K. (1945) 'On problem solving', *Psychological Monographs*, 58, whole, no.270, pp.1–113.
- Ericsson, K.A. (1991) 'Prospects and limits of the empirical study of expertise: an introduction' in Ericsson, K.A. and Smith, J. (eds) *Towards a General Theory of Expertise: Prospects and Limits*, Cambridge, MA, Cambridge University Press.
- Ericsson, K.A. and Harris, M. (1990) 'Expert chess memory without chess knowledge. A training study' Poster presentation at the 31st Meeting of the Psychonomics Society, New Orleans.
- Ericsson, K.A., Krampe, R.T. and Tesch-Rohmer, C. (1993) 'The role of deliberate practice', *Psychological Review*, vol.100, no.3, pp.363–406.
- Ericsson, K.A. and Polson, P.G. (1988) 'Memory for restaurant orders' in Chi, M.T.H., Glaser, R. and Farr, M. (eds) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum.
- Falkenhainer, B., Forbus, K.D. and Gentner, D. (1986) 'The structure-mapping engine', *Proceedings of the Meeting of the American Association for Artificial Intelligence*, pp.272–7.
- Gentner, D. (1983) 'Structure-mapping: A theoretical framework for analogy', *Cognitive Science*, vol.7, pp.155–70.
- Gentner, D., Bowdle, B., Wolff, P. and Boronat, C. (2001) 'Metaphor is like analogy' in Gentner, D., Holyoak, K.J. and Kokinov, B.N. (eds) *The Analogical Mind: Perspectives from Cognitive Science*, Cambridge, MA, MIT Press, pp.199–253.
- Gentner, D. and Gentner, D.R. (1983) 'Flowing waters or teeming crowds: mental models of electricity' in Gentner, D. and Stevens, A.L. (eds) *Mental Models*, Hillsdale, NJ, Lawrence Erlbaum Associates, pp.99–129. (Reprinted in Brosnan, M.J. (ed.) *Cognitive Functions: Classic Readings in Representation and Reasoning*, Eltham, London, Greenwich University Press.)

- Gentner, D. and Markman, A.B. (1997) 'Structure mapping in analogy and similarity', *American Psychologist*, vol.52, no.1, pp.45–56.
- Gick, M.L. and Holyoak, K.J. (1980) 'Analogical problem solving', *Cognitive Psychology*, vol.12, pp.306–55.
- Gilhooly, K.J., McGeorge, P., Hunter, J., Rawles, J.M., Kirby, I.K., Green, C. and Wynn, V. (1997) 'Biomedical knowledge in diagnostic thinking: the case of electrocardiogram (ECG) interpretation', *European Journal of Cognitive Psychology*, vol.9, no.2, pp.199–223.
- Gilhooly, K.J., Phillips, L.H., Wynn, V., Logie, R.H. and Della Sala, S. (1999) 'Planning processes and age in the 5 disk Tower of London task', *Thinking and Reasoning*, vol.5, no.4, pp.339–61.
- Glucksberg, S. and Danks, J.H. (1968) 'Effects of discriminative labels', *Journal of Verbal Learning and Verbal Behaviour*, vol.7, pp.72–6.
- Green, A.J.K. (2002) 'Learning procedures and goal specificity in learning and problem solving tasks', *European Journal of Cognitive Psychology*, vol.14, no.1, pp.105–26.
- Green, A.J.K. and Gilhooly, K.J. (1990) 'Individual differences and effective learning procedures: The case of statistical computing', *International Journal of Man-Machine Studies*, vol.33, pp.97–119.
- Green, A.J.K. and Wright, M.J. (2003) 'Reduction of task-relevant information in skill acquisition', *European Journal of Cognitive Psychology*, vol.15, no.2, pp.267–90.
- Haider, H. and Frensch, P.A. (1996) 'The role of information reduction in skill acquisition', *Cognitive Psychology*, vol.30, no.3, pp.304–37.
- Haider, H. and Frensch, P.A. (1999a) 'Eye movement during skill acquisition: more evidence for the information reduction hypothesis', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.25, no.1, pp.172–90.
- Haider, H. and Frensch, P.A. (1999b) 'Information reduction during skill acquisition: The influence of task instruction', *Journal of Experimental Psychology: Applied*, vol.5, no.2, pp.129–51.
- Hall, T. (1970) *Carl Friedrich Gauss: A Biography* (trans. by A. Froderberg), Cambridge, MA, MIT Press.
- Holding, D.H. (1979) 'The evaluation of chess positions', *Simulation and Games*, vol.10, pp.207–21.
- Holding, D.H. (1985) *The Psychology of Chess Skill*, Hillsdale, NJ, Erlbaum.
- Holding, D.H. and Reynolds, R.I. (1982) 'Recall or evaluation of chess positions as determinants of chess skill', *Memory and Cognition*, vol.10, pp.237–42.
- Holyoak, K.J. and Koh, K. (1987) 'Surface and structural similarity in analogical transfer', *Memory and Cognition*, vol.15, pp.332–40.
- Holyoak, K.J. and Thagard, P. (1989) 'A computational model of analogical problem solving' in Vosniadou, S. and A. Ortony (eds) *Similarity and Analogical Reasoning*, Cambridge, MA, Cambridge University Press, pp.242–66.

- Jeffries, R., Polson, P.G., Razran, L. and Atwood, M.E. (1977) 'A process model for missionaries-cannibals and other river crossing problems', *Cognitive Psychology*, vol.9, pp.412–20.
- Keane, M. (1988) *Analogical Problem Solving*, Chichester, Ellis Horwood.
- Keane, M.T. (1994) 'Constraints on analogical mapping: A comparison of three models', *Cognitive Science*, vol.18, no.3, pp.387–438.
- Knoblich, G., Ohlsson, S., Haider, H. and Rhenius, D. (1999) 'Constraint relaxation and chunk decomposition in insight problem solving', *Journal of Experimental Psychology – Learning, Memory and Cognition*, vol.25, no.6, pp.1543–55.
- Kolers, P.A. (1975) 'Memorial consequences of automatized encoding', *Journal of Experimental Psychology: Human Learning and Memory*, vol.1, pp.689–701.
- Larkin, J., McDermott, J., Simon, D.P. and Simon, H.A. (1980) 'Models of competence in solving physics problems', *Cognitive Science*, vol.4, pp.317–45.
- Lesgold, A.M., Rubinson, H., Feltovich, P.J., Glaser, R., Klopfer, D. and Wang, Y. (1998) 'Expertise in a complex skill: Diagnosing X-ray pictures' in Chi, M.T.H., Glaser, R. and Farr, M. (eds) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum, pp.311–42.
- Logan, G.D. (1988) 'Toward an instance theory of automatization', *Psychological Review*, vol.95, pp.492–527.
- Luchins, A.S. and Luchins, E.H. (1959) *Rigidity of Behaviour*, Eugene, OR, University of Oregon Press.
- Luger, G.F. (1976) 'The use of state-space to record the behavioural effects of subproblems and symmetries on the Tower of Hanoi problem', *International Journal of Man-Machine Studies*, vol.8, pp.411–21.
- McKeithen, K.B., Reitman, J.S., Rueter, H.H. and Hirtle, S.C. (1981) 'Knowledge organization and skill differences in computer programmers', *Cognitive Psychology*, vol.13, pp.307–25.
- Neisser, U., Novick, R. and Lazar, R. (1963) 'Searching for ten targets simultaneously', *Perceptual and Motor Skills*, vol.17, pp.955–61.
- Ohlsson, S. (1992) 'Information processing explanations of insight and related phenomena' in Keane, M.T. and Gilhooly, K.J. (eds) *Advances in the Psychology of Thinking*, London, Harvester Wheatsheaf.
- Renkl, A. (1997) 'Learning from worked-out examples: A study of individual differences', *Cognitive Science*, vol.21, no.1, pp.1–29.
- Schoenfeld, A.H. and Herrmann, D.J. (1982) 'Problem perception and knowledge structure in expert and novice mathematical problem solvers', *Journal of Experimental Psychology – Learning, Memory and Cognition*, vol.8, no.5, pp.484–94.
- Schraagen, J.M. (1993) 'How experts solve a novel problem in experimental design', *Cognitive Science*, vol.17, no.2, pp.285–309.
- Schunn, C.D. and Anderson, J.R. (1999) 'The generality/specificity of expertise in scientific reasoning', *Cognitive Science*, vol.23, no.3, pp.337–70.
- Simon, H.A. and Chase, W.G. (1973) 'Skill in chess', *American Scientist*, vol.61, pp.394–403.

- Simon, H.A. and Hayes, J.R. (1976) 'The understanding process: problem isomorphs', *Cognitive Psychology*, vol.8, pp.165–90.
- Simon, H.A. and Reed, S.K. (1976) 'Modelling strategy shifts on a problem solving task', *Cognitive Psychology*, vol.8, pp.86–97.
- Snoddy, G.S. (1926) 'Learning and stability', *Journal of Applied Psychology*, vol.10, pp.1–36.
- Spellman, B.A. and Holyoak, K.J. (1992) 'If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles', *Journal of Personality and Social Psychology*, vol.62, pp.913–33.
- Sternberg, R.J. (1995) 'Expertise in complex problem solving' in Frensch, P.A. and Funke, J. (eds) *Complex Problem Solving*, Hillsdale, NJ, Erlbaum.
- Sweller, J. (1988) 'Cognitive load during problem solving: effects on learning', *Cognitive Science*, vol.12, no.2, pp.257–85.
- Sweller, J., Mawer, R.F. and Ward, M.R. (1983) 'Development of expertise in mathematical problem solving', *Journal of Experimental Psychology: General*, vol.112, no.4, pp.639–61.
- Thomas, J.C. Jr (1974) 'An analysis of behaviour in the hobbits–orcs problem', *Cognitive Psychology*, vol.6, pp.257–69.
- Thorndyke, P.W. and Stasz, C. (1980) 'Individual differences in procedures for knowledge acquisition from maps', *Cognitive Psychology*, vol.12, pp.137–75.
- Vollmeyer, R., Burns, B.D. and Holyoak, K.J. (1996) 'The impact of goal specificity on strategy use and the acquisition of problem structure', *Cognitive Science*, vol.20, pp.75–100.
- Voss, J.F., Greene, T.R., Post, T.A. and Penner, B.C. (1983) 'Problem solving skill in the social sciences' in Bower, G. (ed.) *The Psychology of Learning and Motivation*, vol.17, New York, Academic Press.
- Zhang, J. and Norman, D.A. (1994) 'Representations in distributed cognitive tasks', *Cognitive Science*, vol.18, no.1, pp.87–122.

# Judgement and decision making

## Chapter 11

Peter Ayton

---

### 1 Introduction

How do people make judgements and decisions? The question has become a steadily increasing preoccupation of cognitive psychology. Plainly, making decisions is a fundamental and everyday human and animal (and, perhaps, machine) activity. Yet, until the 1950s psychologists had hardly given the question any serious thought. Doubtless, this had something to do with the dominance of the behaviourist school of thought throughout the first half of the twentieth century. The behaviourists assumed that human behaviour could be explained entirely in terms of reflexes, stimulus-response associations, and the effects of reinforcers upon them. Accordingly, they shunned the study of mental processes and entirely excluded ‘mental’ terms like desires and goals. As an historical consequence, the foundations of decision research, and hence its contemporary shape, have been strongly influenced by thinking from disciplines *outside* psychology – specifically from mathematics and economics.

This influence from outside psychology left its mark – mathematicians and economists have different concerns to psychologists. The question posed and pursued by thinkers from outside psychology was not how *do* people actually make decisions but how, ideally, *should* decisions be made? What are *good* judgements and decisions and how should we recognize them? As we will see, *behavioural* judgement and decision research – the investigation of *how* people make decisions – has been strongly influenced by a fundamental underlying premise: that the objective of decision making should be to make the ‘best’ choice, and that the best choice can, by some method, be computed.

Judgement and decision making are sometimes distinguished on the basis that *judgements are what underlie decisions*. Judgements can be estimates of some objective quantity – how far away is this object? How dangerous is that hobby? Decisions typically reflect judgements of the qualities of options – but also the preferences of the decision maker.

Of course, real people are not idealized decision-making machines or supercomputers; they do not have unlimited time, knowledge and computational power but a rather limited information-processing capacity. Accordingly, not infrequently, people make mistakes – for example, they may overlook or forget important considerations; they also get bored, suffer anxiety and may not always be sure quite what they want or are trying to achieve. As a consequence, what people do is not always quite the same as what they themselves would agree that they *should* do.

## 1.1 Theories of decision making

Psychologists are of course interested in understanding what people actually do, but this has very often been studied in comparison to what it has been assumed they *should* do. As a result, there are two types of theory of decision making – the ‘ought’ and the ‘is’ – commonly referred to as **normative** and **descriptive** theories respectively. Normative theories define the supposed ideal decision while descriptive theories attempt to characterize how people actually make decisions.

The very existence of this dichotomy suggests that perhaps human decision making is faulty. Indeed, debating whether or not people are essentially rational or irrational decision makers has long been a preoccupation of researchers in this field (cf. Cohen, 1981), just as the rationality of thought has been a key concern for researchers studying human reasoning (as you will see in Chapter 12). However, noting a disparity between the ideal and the actual should not, in itself, cause us to leap to the conclusion that there is something *fundamentally* wrong with the way people make judgements and decisions. In other areas of cognitive psychology, such a step would be seen as clearly absurd; for instance, human memory is manifestly fallible and yet we do not conclude from this that people’s memories are inherently inadequate for the purpose of living their lives.

While persistent errors of judgement or choice could be taken to indicate a fundamental irrationality, researchers in judgement and decision making have tended to adopt a similar position to researchers working in vision. Vision scientists, for instance, do not conclude from the robustness of the Müller-Lyer illusion (see Section 1, Chapter 3) that people are generally poor at inferring object lengths – let alone that visual perception is fundamentally incompetent. Nevertheless, as we will see, people do make judgements and decisions that are inconsistent with normative theory.

## 1.2 Supporting decision making

If people don’t behave as normative theories prescribe, what can be done about it? What should people do to make better choices? What instruction, modes of thinking or decision aids can help real people to make better decisions? A third strand of research straddling the normative and descriptive – the **prescriptive** approach – investigates how to help people make better decisions. One prescriptive approach is decision analysis. Decision analysis is the attempt to help people to make better decisions that conform to normative theory. However, decision analysis is more than just that: helping people to understand and explicate their own objectives and values, search for options and evidence and appreciate their implications is not a straightforward matter. Decision analysis uses a number of techniques, including decision trees (which you will meet in Section 2.1), to help people decompose complex decisions into more manageable components, elicit values and beliefs for the elements and apply normative principles to their reintegration.

## Summary of Section 1

---

- Judgements underlie decisions.
- Researchers distinguish between actual (descriptive) and ideal (normative) decision making.
- Decision analysis can support people in making better decisions – a prescriptive approach.

## 2 Normative theory of choice under risk

In many situations where we must make a choice, we will be uncertain about whether the possible outcomes will turn out to be good or not so good. Consequently, risk is an inescapable fact of life. Some sorts of risky decisions are easy to imagine: a person may have to consider whether to continue living with a debilitating health condition or risk surgery that might help but could leave them worse off. Investment decisions often involve contemplating whether to put money in a safe investment with a small return or a riskier investment that might yield a lot of money but could lose everything. Decisions of this sort can be analysed as **gambles**. Gambling is the dominant metaphor in decision research as gambles involve *uncertainty* about what will happen.

The most extensively applied normative model of risky choice is called **subjective expected utility** theory (SEU). This theory is an extension by Savage (1954) of the ‘expected utility’ theory published by von Neumann and Morgenstern (1944) in their book *Theory of Games and Economic Behavior*. Von Neumann and Morgenstern’s analysis was applied to games of chance with known or computable probabilities. Savage’s extension of the theory allowed for what he called ‘personal probabilities’ – commonly referred to nowadays as ‘subjective probabilities’. Savage’s generalization of the theory allows it to be applied to decision situations where no objective mathematical probabilities are available and where judgements may be no more than expressed beliefs about likelihoods.

For example: imagine contemplating an invitation to a picnic. Suppose you have to write an essay over the weekend in question but do not want to miss out on anything really good. On the other hand, you wouldn’t want to waste time at a boring or horrible picnic. You may be unsure whether it will rain or not; whether Tarquin (an individual about whom you have very strong views) will be present or not; and there could be any number of other factors that would affect the value of accepting the invitation. So how should you decide?

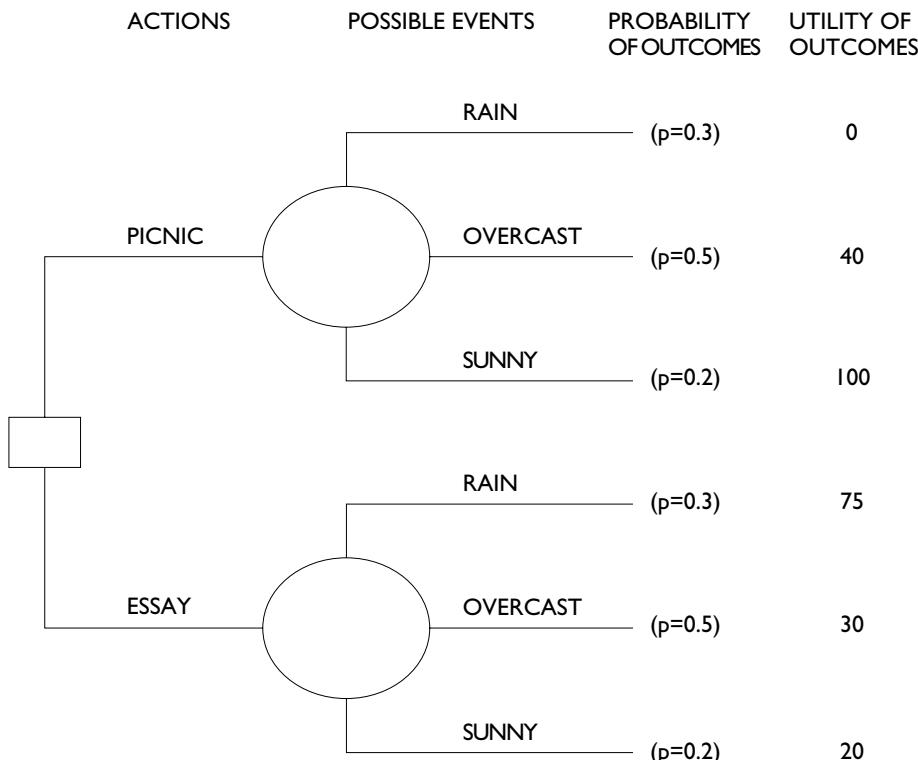
According to standard normative theory, a rational decision maker should trade off the value of all the possible outcomes by the likelihood of obtaining them. Just as the value of a lottery ticket will vary according to both the value of the prizes *and* the chances of winning them, so, according to the normative theory of choice, do choice options. What is the value of an idyllic or dreadful picnic? What is the value of any of the alternative activities you could indulge in? How likely is it to rain or that Tarquin will be there? All the relevant elements must be quantified and combined to compute

the optimal decision. How this is done is illustrated in the next section where the technique of decision analysis is described.

## 2.1 Prescriptive application of normative theory: decision analysis

**Decision analysis** is a technology based on SEU that was developed in the 1960s to improve decision making (cf. Raiffa, 1968; Schlaifer, 1969). Decision analysts use normative theory to represent decision problems so that, ideally at least, the normatively correct decision can be computed. In the classic decision analytic framework (von Winterfeldt and Edwards, 1986) numerical probabilities are assigned to all the different events identified in a **decision tree**. The decision tree is simply a means of representing or modelling the decision. The best alternative is then selected by combining the probabilities and the utilities corresponding to the possible outcomes associated with each of the possible alternatives.

Figure 11.1 shows a simple decision tree for our student trying to decide whether to go on a picnic or stay at home and write an essay. The tree portrays two possible actions or events – picnic or essay – and three possible future events (weather conditions) that would affect the value or utility of the resulting six identifiable outcomes. Of course, there could be many more options (students' weekend options involve more than essays and picnics) and possible future conditions (there may well be more than the weather to consider). The utilities, as in this case, will often reflect subjective evaluations of the quality of the outcomes – though for business or



**Figure 11.1** Decision tree for deciding whether to go on a picnic or write an essay

financial decisions it might reflect monetary profits. Here, the scale for utility is arbitrary – let's suppose the student was asked to rate each of the six possible outcomes on a 0–100 scale assigning 0 to the worst outcome, 100 to the best and scaling the others relative to those. A picnic in the sun is rated the best outcome and a picnic in the rain the worst. Writing an essay at home is affected by the weather, albeit differently, because, as any student knows, a sunny day is the worst time to have to stay in and work – especially if you know your friends are out having fun. These figures might not reflect your own utilities for these outcomes – utility is personal. You might revel in rainy picnics – if so, your utilities for this decision would be quite different.

We also need likelihoods for the three different weather conditions. In some countries weather forecasters routinely broadcast these, but as long as they accurately reflect our real beliefs, we could use our own judgements. Decision theory can only tell you how to decide given your beliefs about the utility and likelihood of the relevant events. With these data, we can now compute the expected utility of the two actions identified in the tree. The expected utility of each *outcome* is defined as the probability or likelihood of that outcome ( $P$ ) multiplied by its utility ( $U$ ). The expected utility of an *action* is the sum of such expected utilities for each of the possible *outcomes*. So, the expected utility of the picnic *action* is the sum of the expected utilities for the rainy picnic, the overcast picnic and the sunny picnic:

$$\begin{aligned} SEU(\text{picnic}) &= \sum P \times U = P(\text{rainy picnic}) \times U(\text{rainy picnic}) \\ &\quad + P(\text{overcast picnic}) \times U(\text{overcast picnic}) \\ &\quad + P(\text{sunny picnic}) \times U(\text{sunny picnic}) \\ &= 0.3 \times 0 + 0.5 \times 40 + 0.2 \times 100 \\ &= 40 \end{aligned}$$

That is, by multiplying the utility of each outcome by the likelihood of obtaining it, we can calculate that the expected utility for the picnic is 40. Similarly, we can calculate that the expected utility for the essay is 41.5. Because the expected utility of the essay is greater than that for the picnic *this* student should write the essay.

The difference between 41.5 and 40 may seem rather close, but remember that the numerical scales are arbitrary. In practice, a decision analyst using this procedure as part of the *prescriptive* approach to assist decision making would go back and check that reasonable variations in the values entered in the tree do not alter the decision. If they do, then the user must be sure that the numbers genuinely reflect their beliefs and values.

## ACTIVITY 11.1

---

Try to produce a decision tree for two activities you might consider doing. For example, you might want to compare going to a party with going to the cinema. Your enjoyment of a party may depend on who else is going, and where the party is. Going to the cinema may be fun if you can see a film you particularly want to

see, or may be just an opportunity to while away some time if you are indifferent to the films on show. Think about which you would choose to do. Now try computing the expected utility for the two actions. Which action does the analysis suggest you *should* choose? Are they the same and, if not, why do you think they differ?

### COMMENT

The calculations illustrate how the expected utility theory defines a normative decision, but why follow this procedure? The assumption is that you might not make such a good decision by relying on unaided intuition. The decision tree can help you to clarify the relevant events and the structure of the decision while the computations based on your stated values will follow the normative theory. Alas, there is no normative technique for eliciting the correct structure of decisions from individuals. However, in decision-analytic practice decision trees are used more to facilitate thought rather than to definitively represent complex decisions; a realistic portrayal of all relevant act–event combinations might result in a complicated mess.

---

Since its initial inception, the role of decision analysis has changed (Phillips, 1989). Nowadays, decision analysts view decision trees as tools to aid thinking, not as providing solutions (von Winterfeldt and Edwards, 1986). The theory of **requisite decision modelling** (Phillips, 1984) claims that models of decisions need only be sufficient in structure and content to resolve the issues at hand. A good model captures the essential elements of a decision situation for the purposes of the decisions to be made. An iterative procedure is followed involving constructing the model, analysis, model refinement, and subsequent re-analysis. At the point when no additional intuitions emerge from further analysis, the model is said to be *requisite*. The claim is that this procedure helps to develop a shared understanding and fosters commitment to the way forward. Note that these are social purposes. The technical computations of a decision analysis are less than half the story of improving decision making. Constructing representations of decisions and eliciting values are not achieved by mechanical operations; they often involve deep thought – or what some (cf. Watson and Buede, 1987) call ‘soul searching’. Perhaps the greatest virtue of decision analysis is that it obliges decision makers to make explicit all the bases of a decision.

In the next section, we shall examine some of the principles underlying SEU, and consider the extent to which SEU accurately describes human decision making.

## 2.2 Axioms underlying subjective expected utility theory

Mathematicians such as von Neumann and Savage established that SEU is implied by the acceptance of certain principles or axioms: **comparability**, **transitivity**, **dominance**, **independence** and **invariance**. According to SEU, if a decision maker violates one or more of these axioms, then their choices will not maximize expected utility and so will not be normative. The axioms therefore define a kind of coherence

to our choices and give them internal consistency. We shall look at them in more detail now:

- Comparability (or completeness)

If you have to evaluate two alternatives A and B you must be able to say whether you:

- 1 prefer A to B or
- 2 prefer B to A or
- 3 are indifferent between A and B.

- Transitivity

If you prefer A to B and B to C then you must prefer A to C. That is, choices should be capable of being ordered.

- Dominance

An option is dominant and must be preferred if, when compared to another option, it is better in at least one respect and at least as good or better in every other respect. *Dominated* options must never be preferred.

- Independence

If there is some outcome that is unaffected by, or independent of, your choice then this outcome should not affect your choice.

- Invariance

Different representations of the same choice problem should result in the same choices. That is, the preference for options should be independent of how they are described.

### 2.3 Violations of the axioms

Although the axioms might strike you as uncontentious and straightforward, they can be questioned. For example, the comparability axiom is threatened by claims that people may not be indifferent just because they are unable to say which of two states they prefer. Curiously, even the original architects of the theory admitted that it might not always be a reasonable assumption:

We have conceded that one may doubt whether a person can always decide which of two alternatives ... he prefers. If the general comparability assumption is not made, a mathematical theory ... is still possible

*(von Neumann and Morgenstern, 1944, pp.19–20)*

The axioms also seem to vary somewhat in their intuitive appeal; while independence and transitivity might not be obvious requirements for rational choice, dominance and invariance appear essential. Nonetheless, psychologists have shown that, under certain conditions, systematic violations of each of the axioms can be observed in people's choices (Tversky and Kahneman, 1986). Since violations of the axioms imply that people are not choosing according to the normative theory, we

could conclude one of two things: either that there is something wrong with the choices or that there is something wrong with the normative theory (or perhaps both) – in any case, as we shall now see, SEU does not provide a good description of actual human choices.

### 2.3.1 Violations of transitivity

Observed violations of the transitivity axiom have generally led to the conclusion that people's choices are not ideal. For example, Tversky (1969) asked people to state their preferences between pairs of college applicants rated on three dimensions – intelligence, emotional stability and social facility (as in Table 11.1).

**Table 11.1** Ratings of five applicants on three dimensions

Applicant	Intelligence	Emotional stability	Social facility
A	69	84	75
B	72	78	65
C	75	72	55
D	78	66	45
E	81	60	35

Now try Activity 11.2.

### ACTIVITY 11.2

Consider the following pairs of applicants in Table 11.1: A–B, B–C, C–D, D–E and E–A. For each pair, write down which of the two applicants you would prefer given their ratings. Do note, however, that you should weight intelligence more highly than either of the other dimensions.

### COMMENT

Tversky's subjects were presented with all possible pairs of applicants (together with some others), one pair at a time and were similarly told to weight intelligence more than the other two dimensions. Subjects typically preferred A to B; B to C; C to D; and D to E. However, violations of transitivity were demonstrated by the typical simultaneous preference for E over A. Did your own preferences coincide with these? If not, try to work out why Tversky's subjects might have adopted the preferences they did.

If people reliably mapped all the dimension scores of each option onto a common currency of utility, then *systematic* violations of transitivity would not occur – so demonstrations of intransitive preference are revealing about the nature of the choice process. For Tversky (1969, p.46), this was key: 'The main interest in the present results lies not so much in the fact that transitivity can be violated but rather in what these violations reveal about the choice mechanism.' Tversky suggested ways in which decision making might be rendered less cognitively demanding by applying decision rules that simplify the task. He offered two hypotheses about the choice

process: (1) people compare the alternatives on each dimension in turn, rather than *evaluating* each option on all dimensions before comparing overall evaluations, and (2) that people ignore dimensions on which the alternatives – even if discriminable – are rated similarly.

For example, when comparing successive pairs in the chain such as A and B on intelligence, subjects may decide that the difference between them is negligible – and so, in the interests of simplifying the decision, ignore it altogether. However, small differences add up – at the ends of the chain the difference in intelligence between A and E is too big to ignore – hence, the observed pattern of intransitivity. Note that this explanation (that people try to simplify decisions by ignoring information) assumes that people have limited information-processing capacity.

In relation to SEU, intransitivity is an irrational pattern of choice but it may be reassuring to note that it is not a uniquely human condition. For instance, in an experiment where bees chose between artificial flowers that offered varying amounts of nectar with varying degrees of accessibility, Shafir (1994) found that they violated transitivity in their foraging preferences for flowers. As bees have been successfully foraging for millions of years it is tempting to assume that perhaps the costs – intransitive preferences cannot maximize expected utility – are outweighed by the gains – presumably, reduced information processing.

When confronted with evidence of intransitivity in their choices, people typically immediately concede that there is some inconsistency and are usually willing to change their choices to preserve transitivity. Hence, they seem to endorse the normative status of the axiom even though their violations show that transitivity is not descriptive of human choice.

### 2.3.2 Violations of the independence axiom

Violations of the independence axiom are more problematic, and have proved a serious challenge to both the normative and descriptive status of SEU. The first challenge came from the French economist and Nobel laureate Maurice Allais who published a paper in 1953 describing what is now called the **Allais paradox** (Allais, 1953; 1979).

Allais observed that people are reluctant to exchange a certain prospect of something wonderful (e.g. receiving \$1,000,000) for a not quite certain prospect of something even more wonderful (e.g. 99 per cent chance of receiving \$5,000,000). The paradox occurs because if both the above prospects are reduced in likelihood by a similar amount (so that neither offers certainty) people *are* usually willing to exchange a smidgen of likelihood for a substantial increase in benefit. Box 11.1 shows how this is a problem for SEU.

Allais made it perfectly clear that he considered that the intuitions which produced the paradox *should* over-rule the independence axiom, that is, the normative theory was not valid. He even claimed: ‘It is quite disappointing to have to exert so much effort to prove the illusory character of a formulation whose oversimplification is evident to anyone with a little psychological intuition’ (Allais and Hagen, 1979, p.105). Others, including Savage, who, embarrassingly, initially succumbed to the paradox in his own choices, felt differently and argued that the intuitions underlying the choices were wrong and that the theory was normatively correct.

## 11.1

### The Allais paradox

**Table 11.2** The Allais paradox as a choice of lotteries: each lottery involves 100 tickets. The table shows the number of tickets that win anything from \$0 to \$5,000,000

		<b>Lottery ticket numbers (1–100)</b>		
		<b>I</b>	<b>2–11</b>	<b>12–100</b>
Situation 1	Choice A	\$1,000,000	\$1,000,000	\$1,000,000
	Choice B	\$0	\$5,000,000	\$1,000,000
Situation 2	Choice C	\$1,000,000	\$1,000,000	\$0
	Choice D	\$0	\$5,000,000	\$0

Table 11.2 shows two separate situations where you can choose to take part in one of two lotteries and draw one ticket from the lottery you choose. In Situation 1 you can choose between lotteries A and B. If you are like most people you would choose A, as this guarantees \$1,000,000. B could deliver \$5,000,000 but there is a small chance of ending up with nothing at all.

However, when faced with the choice in Situation 2, between C and D, most people prefer D – now they are willing to face a very slight increase in the prospect of getting nothing at all in order to have a chance of winning \$5,000,000. To see how this violates the independence axiom, simply cover up the last column – now the two situations appear identical. As the contents of the last column are identical (\$1,000,000) for A and B in Situation 1, and also for C and D (\$0) in Situation 2, then, according to the axiom, the information in this column should not influence your choice. So, if you prefer A to B, you should also prefer C to D.

When made aware that they are violating the independence axiom, people sometimes alter their choices to conform to it (Keller, 1985) but sometimes – even after a thorough explanation of its virtues – they don't (Slovic and Tversky, 1974). Slovic and Tversky suggested that people may alter their choices to concur with the axiom – not through appreciating the merits of so doing – but because they might be intimidated by the suggestion that not doing so would be irrational. Their paper concludes with a delightful imaginary debate between Savage and Allais wherein Savage insists that people only reject the axiom when they do not understand it, while Allais (who plausibly claimed to both understand and reject the axiom) asks how Savage could distinguish between failure to understand the axiom and enlightened rejection of it. The debate highlights an irresolvable conflict between two different intuitions – those that support the axiom and those that support the pattern of choices in the Allais paradox. Ultimately, rather like the Ten Commandments (which are also often violated), the normative status of SEU and its axioms is not in any sense a demonstrable truth – they only appeal (or not) as principles to live by.

## Summary of Section 2

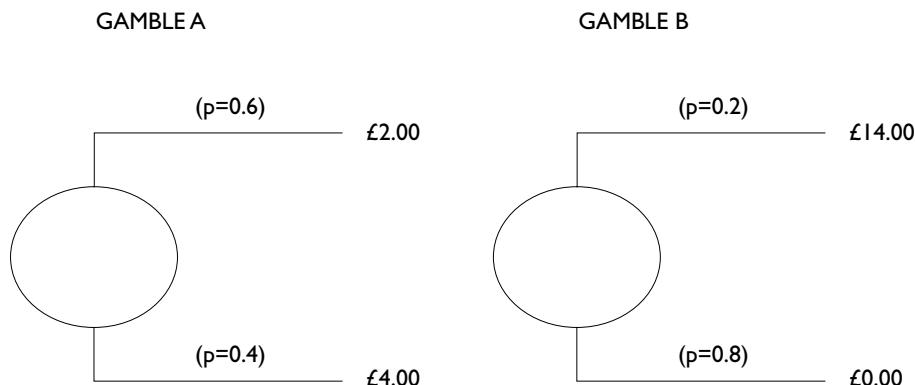
---

- SEU provides a normative theory of decision making under uncertainty.
- Decision analysis offers a prescription for making decisions using SEU.
- Conforming to SEU is equivalent to adhering to certain axioms.
- Human decision making has been shown to violate these axioms, implying that it is not adequately described by SEU.

### 3 Findings from behavioural decision research

Most decision researchers accept the normative status of SEU but also consider that it does not describe human decision making. Some thirty years after the emergence of SEU, Slovic *et al.* (1977, p.9) reviewed the psychological literature and commented: ‘... during the past 5 years, the proponents of SEU have been greatly outnumbered by its critics’. Edwards (1992) polled an all-star cast of leading decision theorists at a conference. They unanimously endorsed traditional SEU as the appropriate normative model but unanimously agreed that people don’t behave as the model requires. Nonetheless, and perhaps in spite of the survival of SEU as a normative theory (albeit on the basis of opinion polls), Allais was awarded the Nobel Prize for Economics in 1988.

Violations of the axioms of SEU imply that it does not provide a valid description of human decision making. There is now a considerable mass of empirical evidence indicating that SEU does not predict human decisions either. One piece of evidence comes from Edwards (1955), who offered experimental subjects choices between bets of equal expected value such as the choice between Gambles A and B in Figure 11.2 below. If you accept Gamble A, it will give a 0.6 probability (or 60 per cent chance) of winning £2.00, and a 0.4 chance of £4.00; Gamble B gives a 0.2 probability of winning £14.00 and a 0.8 probability of winning nothing.



**Figure 11.2** Two simple gambles of equal expected value

If we assume that the amounts of money are in direct proportion to people's utility for each outcome, then we can calculate the expected utility (EU) for the bets:

$$EU(\text{Gamble A}) = (0.6 \times £2) + (0.4 \times £4) = 2.8$$

$$EU(\text{Gamble B}) = (0.2 \times £14) + (0.8 \times £0) = 2.8$$

In a series of experiments, Edwards found that most people have definite preferences between bets of equal expected value. Compared to a good chance of winning a small amount they preferred a long shot of winning a large amount – provided there was no chance of losing very much. People strongly avoided gambles that involved even a low probability of losing a lot. Edwards concluded that SEU was not a guide for their choices between gambles. Later, Lichtenstein *et al.* (1969) found that expected value is irrelevant even when the concept was carefully explained to respondents.

### 3.1 The 'preference reversal phenomenon'

Far worse was to come for SEU however when the **preference reversal phenomenon** was discovered. Slovic and Lichtenstein (1968) had noticed that ratings of a gamble's attractiveness, as well as choices between pairs of gambles, were strongly influenced by the *probability* of winning and losing. Yet when asked how much they would be willing to pay in order to take the gamble, or the smallest amount they would be willing to sell the gamble for, people were more heavily influenced by the *amounts* that could be won or lost. Lichtenstein and Slovic (1973) realized that if there was a different basis for *choosing* than for *valuing* it should be possible to construct pairs of gambles so that people would prefer A to B but pay more for B than A. They were able to demonstrate this effect in a series of studies – including one conducted with real gambles in a Las Vegas casino. Typically, one bet would have a high probability of winning a modest amount (called the 'p bet') while the other would offer a lower probability of winning a higher amount (called the '\$ bet'):

*p bet : 11/12 chance of winning 12 chips*

*1/12 chance of winning 24 chips*

*\$ bet : 2/12 chance of winning 79 chips*

*10/12 chance of losing 5 chips*

These two gambles were chosen equally often by the casino subjects; however, the \$ bet received a higher selling price about 88 per cent of the time. Among those choosing the p bet, 87 per cent gave a higher selling price to the \$ bet. So, people value the \$ bet more highly than the *p* bet, but don't prefer the \$ bet any more than the *p* bet. From a rational perspective this is a hopeless pattern of behaviour.

The finding, replicated numerous times since (Slovic, 1995), clearly poses a major threat to the SEU view of rational choice. Two economists, Grether and Plott (1979, p.623), realizing that: 'it suggests that no optimisation principle of any sort lies behind even the simplest of human choices', conducted a series of studies 'designed to discredit the psychologists' works as applied to economics'. However,

even after controlling for all the economic explanations of the phenomenon that they could find – including that the experiment be conducted by economists rather than psychologists ('Psychologists have the reputation for deceiving subjects', p.629) – the reversals persisted.

### 3.2 Causes of anomalies in choice

Why do preference reversals occur? Slovic (1995) summarized the evidence in favour of a 'scale compatibility hypothesis'. The idea is that the weight of an option attribute in judgement or choice is influenced by its compatibility with the response mode. As economic value is expressed in terms of money, subjects find it easier to use the monetary aspect of the gamble to set the *value* of the gamble. However, when asked which gamble they *prefer* subjects have no similarly compelling reason to weight the monetary aspect of the gamble to determine their choice. You should note that this explanation does not depend on the presence of risk or uncertainty and indeed Tversky *et al.* (1990) have demonstrated preference reversals for options where no risk is present.

#### 3.2.1 The 'prominence effect'

To account for another aspect of preference reversals, Tversky *et al.* (1988) identified a specific instance of the compatibility effect, which they termed the **prominence effect**. Slovic (1975) had observed that, after earlier adjusting the pay-offs of two gambles so as to make them equally valuable, people did not randomly choose between them but typically chose the gamble with the higher likelihood of winning. Tversky *et al.* (1988) suggested that the more prominent (or important) attribute would weigh more heavily in making a choice than in a matching task (as explained below). For example, in one problem, respondents were asked to imagine two programmes being considered by a transport ministry for dealing with traffic accidents in a country where 600 people are killed every year. Both programmes were described in terms of their annual costs and the expected annual number of casualties that would result if each was introduced. For the choice task, people were asked to choose between the following two options:

Choice task		
Option A	570 casualties	Cost \$12 million
Option B	500 casualties	Cost \$55 million

Of those who took part in the experiment, 67 per cent preferred B to A – note that this implies that the difference in casualties (70) is more important than the difference in costs (\$43 million).

Other respondents performed a matching task, where they had to fill in the missing value so as to make the two programmes equally attractive:

Matching task		
Option A	570 casualties	Cost \$12 million
Option B	500 casualties	Cost \$?

The typical matching value was less than \$55 million – indeed, only 4 per cent of respondents gave a value higher than \$55 million.

Plainly, the trade-off between attributes is different with matching than with choice. Why? Tversky *et al.* argued that *choice* invites more *qualitative* reasoning – people select the option that is superior on the most important attribute (lives saved). This is cognitively simpler, easier to justify and resolves the conflict between the two attributes – albeit by effectively ignoring it. Matching however entails a more quantitative assessment. The matching task cannot be performed at all without paying attention to the values of both attributes and their relative importance.

Real world choices often resemble matching or choice tasks. For example, you might ask yourself what is the most you are prepared to pay when shopping for a particular item (as in the matching task), or you could ask yourself whether you are willing to pay the advertised price for the item (as in the choice task). The evidence suggests that the decisions will tend to diverge. Similar effects may well affect budget setting and resource-allocation decisions. Comparing budget allocation (matching) with budget cutting (choice) the prominence hypothesis suggests that, when forced to choose what items to cut from a hospital budget, health provision (the most important attribute) may fare better than (say) staff pay.

### 3.2.2 Choosing and rejecting options

Shafir (1993) has shown that choosing one of two items is not the complement of rejecting one of the two items. Sometimes when deciding between two options, people both select and reject the same option. When we are trying to select an option we tend to focus on positive features and when we are looking for reasons to reject an item we tend to focus on negative. Thus, items that have obvious positive features will be selected over items that do not. Similarly, items that have obvious negative features will be rejected before items that do not. It seems that rather than rank order options, as mandated by SEU, people look for *reasons* for their decisions. This has led to the proposal of a **reason-based theory of choice** (Shafir *et al.*, 1993) according to which reasons for choosing are more influential when we choose rather than reject, and reasons for rejecting are more influential when we reject rather than choose.

Failure to resolve conflict in choices can also be revealing of reasoning. The economist, Thomas Schelling, tells of an occasion when he went to buy an encyclopaedia for his children (cf. Shafir *et al.*, 1993). At the bookshop he was presented with two attractive encyclopaedias and, finding it difficult to choose between them, went home with neither – despite feeling that he would have happily bought either if it had been the only one available. Unresolved conflict can cause people to defer choosing because they lack a clear reason to select either option.

### 3.2.3 The ‘evaluability principle’

Difficulty in interpreting value is addressed in a study conducted by Hsee (1998) who has developed the notion of **evaluability** to explain a type of preference reversal that occurs when items are evaluated separately or jointly. For example, if shopping for a piano in a musical instrument shop you might compare several pianos. At an auction or second-hand shop, however, you might have to consider a single piano.

Hsee argues that attributes vary in how easy or difficult they are to evaluate, and that their evaluability varies according to whether options are considered in isolation or in relation to other options. In one experiment Hsee asked people to assume they were music students looking for a used music dictionary. In the joint-evaluation condition, participants were shown two dictionaries, A and B (see Table 11.3), and asked how much they would be willing to pay for each. Willingness-to-pay was higher for Dictionary B (\$27) than A (\$19), presumably because of its greater number of entries. However, in the single evaluation condition when one group of participants evaluated only A and another group evaluated only B, the mean willingness to pay was higher for A (\$24) than B (\$20).

**Table 11.3** Attributes of two dictionaries in Hsee's (1998) study

Year of publication	Number of entries	Any defects?
Dictionary A	10,000	No, it's like new
Dictionary B	20,000	Yes, the cover is torn; otherwise it's like new

Hsee explains this reversal by means of the **evaluability principle**. He argues that, without a direct comparison, the 'number of entries' attribute is hard to evaluate because the evaluator does not have a precise notion of *how good* or *how bad* 10,000 (or 20,000) entries is. However, the 'defects' attribute is evaluable because it translates easily into a precise good/bad response – most people find a defective dictionary unattractive and a like-new one attractive – and thus it carries more weight in the independent evaluation. Under joint evaluation, however, the buyer can see that B is far superior on the more important attribute, number of entries. Thus, the 'number of entries' attribute becomes *evaluable* through the comparison process.

## Summary of Section 3

---

- Preference reversals illustrate that SEU fails to predict aspects of human decision making.
- The prominence and evaluability of attributes, and whether a task involves choosing or rejecting, have been shown to influence people's choices, effects not predicted by SEU.

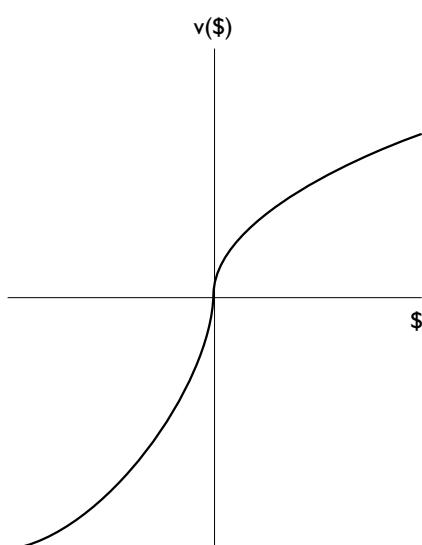
## 4 Prospect theory

One important general conclusion that follows from these demonstrations of anomalies in choice is that people don't have a set of pre-existing stable values, that is, preferences, that they simply apply to choice situations. What is evident is that decisions change because the underlying bases of decisions change according to the demands of the decision task and the nature and context of the information presented. The unstable nature of preferences raises difficult – perhaps even unsolvable – questions regarding people's preferences. If different procedures for eliciting preferences elicit different choices, then how can preferences be defined and how should they be measured?

Kahneman and Tversky (1979) have proposed a descriptive model for decision making under risk, called **prospect theory**, which explains many of the phenomena that cannot be accounted for by SEU. Unlike SEU, prospect theory does not define ideal choices. It is a descriptive, not a normative, theory intended to account for human choices. Prospect theory is essentially an adapted version of SEU, which is modified so as to account for the observed discrepancies with SEU. Prospect theory identifies two phases to the choice process:

- 1 In the editing phase, the decision problem is represented; ‘negligible’ components may be discarded and a reference point is used to enable decision outcomes to be construed as ‘gains’ or ‘losses’.
- 2 In the second phase, attitudes towards risks involving gains and losses are used to evaluate the identified prospects.

Prospect theory proposes that people evaluate decision outcomes in terms of gains or losses from a neutral reference point. Figure 11.3 shows how people are thought to value gains and losses. The horizontal axis to the right of the origin shows objective gains(\$); as they increase the subjective value of the gains( $v(\$)$ ) also increase but with a diminishing slope. This illustrates the fact that, for example, the psychological difference between \$0 and \$10 is greater than that between \$100 and \$110. Notice there is a similar effect for losses – the slope to the left of the origin shows that losses also diminish in a similar fashion. As the slope is not uniform, your *attitude* to risks varies as a function of where you see yourself on the curve. For contemplating gains (to the right of the origin) decisions will tend to be *risk averse* – most people decline to risk a gain of \$10 for a 50 per cent chance of winning \$20. By contrast, with losses, to the left of the origin, decisions tend to be *risk seeking* – in order to avoid a sure loss of \$10 most people would be tempted to risk a 50 per cent chance of losing \$20.

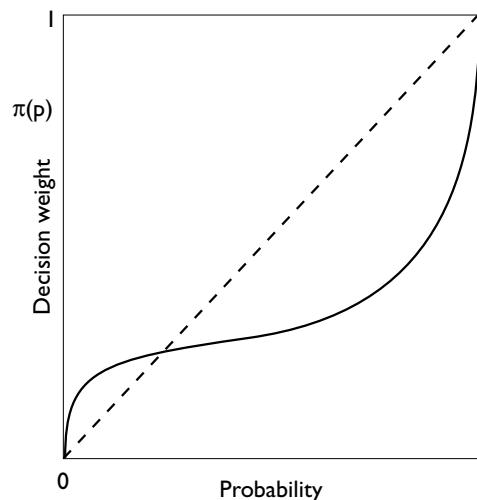


**Figure 11.3** The value function in prospect theory showing the subjective value of a gain or loss as a function of the actual or objective amount of gain or loss

## 4.1 Prospect theory and ‘loss aversion’

Another critical feature of the value function is that the curve is steeper for losses than for gains. This models the observation of **loss aversion** – that people feel losses more than they do gains of equivalent value. Famously, the economist Paul Samuelson once offered a bet to an economist colleague. They would flip a coin and if the colleague won he would get \$200; if he lost he would have to pay Samuelson \$100. The colleague, claiming he would feel the \$100 loss more than the \$200 gain, turned the bet down but mentioned that if Samuelson would play the bet 100 times he would play. (You might have noticed that this pair of preferences is paradoxical with respect to SEU – anyone declining one gamble should not accept any number of plays of the same gamble; Samuelson, 1963.)

Another attitude applied in the evaluation phase is that probability is distorted. Probabilities ( $p$ ) are replaced by decision weights ( $\pi(p)$ ). Note that this distortion does not apply to the judgement or estimation of probability but to the probability that results from judgement or even one supplied to the decision maker. Figure 11.4 shows that low probabilities (except zero, which is given zero weight) are over-weighted. Note the lower end of the curve is above the diagonal dotted line. Moderate and high probabilities (except certainty, which is given the correct weight of 1) are under-weighted: note the upper end of the curve is below the diagonal dotted line.



**Figure 11.4** The weighting function in prospect theory showing the weighted probability  $\pi(p)$  given to probabilities  $p$  varying between 0 and 1

The weighting function of prospect theory accounts for the behaviour observed in the Allais paradox. Because people weight probabilities just below certainty far less than they should, they correspondingly give certainty a *relatively* very high value: 100 per cent certainty is weighted a great deal more than 99 per cent. Moreover, very small probabilities are over-weighted – referring back to Allais we can see that people would worry disproportionately about the 1 per cent chance of not winning.

## 4.2 ‘Framing’ effects

Kahneman and Tversky have reported numerous experiments demonstrating phenomena not anticipated by SEU but predicted by prospect theory. For example, the idea that gains and losses are determined by application of a reference point predicts **framing effects**. Tversky and Kahneman (1981) asked respondents to imagine that the USA was preparing for the outbreak of an unusual disease expected to kill 600 people. Two alternative programmes had been proposed to combat the disease.

- If Programme A is adopted, 200 people will be saved.
- If Programme B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds chance that no people will be saved.

You should note that the options are described in terms of gains – the number of lives that might be saved. Of the respondents, 72 per cent chose Programme A and 28 per cent chose Programme B – definitely saving 200 lives is seen as more attractive than a one-third chance of saving 600 lives. For gains, as we saw in Figure 11.3, people are risk averse – as a result, gains that are certain are more attractive than a gamble of equal expected value.

A second group of respondents was presented with a different description of the two programmes.

- If Programme C is adopted, 400 people will die.
- If Programme D is adopted, there is a one-third probability that nobody will die and a two-thirds chance that 600 people will die.

In this case, only 22 per cent of respondents chose Programme C while 78 per cent chose Programme D. Of course, Programmes C and D are identical to Programmes A and B except that now the outcomes are ‘framed’ in terms of the numbers of lives that might be lost. Framed as a loss, the same risky option becomes more popular than the riskless option (a clear violation of the invariance axiom that you met earlier). The reversal of preference can be explained by the change of the reference point in conjunction with the shape of the value function. With gains, the reference point is defined by what will happen if nothing is done: 600 dead. Programme A looks attractive as it definitely saves 200 while Programme B risks a two-thirds chance of saving nobody. The relative overweighting of certainty will also contribute to the relative attractiveness of the sure gain of Programme A. In the domain of loss, the reference point is defined by the present: nobody has yet died. Programme D is more attractive as 600 deaths are not substantially worse than 400, and it offers a chance that nobody will die.

## Summary of Section 4

---

- Prospect theory can account for an enormous range of observed anomalies in choice both in laboratory experiments and in field data representing real-life decisions.
- People feel losses more than they do gains of equivalent value.
- Framing of options influences preference patterns.

## 5 Judgement under uncertainty

As I noted earlier, judgement and choice can be distinguished on the basis that judgements are what underlie choices. To compute the ideal choice, the normative theory requires inputs of value and likelihood, which typically will be judgements by decision makers. Plainly, evaluating likelihood is a crucial prerequisite for effective decision making under uncertainty. For this reason, research on judgement of likelihood has a particular significance.

### 5.1 Judging probabilities and Bayes' Theorem

In the 1960s, Ward Edwards and his colleagues conducted a number of studies using what were called the **book bag** and **poker chip paradigms**. A typical experiment would involve two opaque bags. Each bag contained 100 coloured poker-chips in different but stated proportions of red to blue. Suppose Bag A contains seventy red and thirty blue chips while Bag B contains thirty red and seventy blue chips. The experimenter first chooses one bag at random and then draws a series of chips from it. After each draw, the poker chip is replaced and the bag well shaken before the next chip is drawn. The subjects' task is to say how confident they are – in probability terms – that the chosen bag is A, containing predominantly red chips, or B, containing predominantly blue chips.

Bayes' Theorem can be used to calculate how the objective probabilities change after each piece of new information, and so can be used to evaluate human performance (see Box 11.2). Where the two competing hypotheses are that the bag is A and the bag is B, and the information is the drawing of a red chip, Bayes' Theorem gives the following equality:

$$\frac{p(A | RED)}{p(B | RED)} = \frac{p(RED | A) \times p(A)}{p(RED | B) \times p(B)}$$

where  $p(A | RED)$  stands for the probability that the bag is A, given that a red chip has been drawn.

Prior to drawing any chips the probability that the bag is A and that the bag is B are both 0.5; if we draw one red chip the likelihood of this is 0.7 for Bag A and 0.3 for Bag B. We can substitute these values into Bayes' Theorem:

$$\frac{p(A | RED)}{p(B | RED)} = \frac{0.7}{0.3} \times \frac{0.5}{0.5}$$

$$\frac{p(A | RED)}{p(B | RED)} = \frac{0.35}{0.15}$$

Therefore, the odds for Bag A over Bag B after drawing one red chip are 0.35: 0.15 or, converting these into percentage probabilities, 70 per cent for Bag A and 30 per cent for Bag B. Bayes' Theorem can then be used after each subsequent drawing of a chip to calculate how these probabilities should change.

Thus, for example, following the drawing of the red chip and prior to drawing a second chip our new prior odds ratio is  $\frac{7}{0.3}$ . If we replaced the red chip and shook the bag so the probabilities of drawing a red and blue chip remained as before, the impact

## 11.2

### Research study

#### Using Bayes' Theorem

So, how good are people at judging probabilities? In order to work this out, we can try using an objective standard to which we can compare human performance. One early benchmark used for this purpose was Bayes' Theorem (which you will also meet in Chapter 12), a mathematical formula used for combining probabilities; it is of interest to decision makers as it can be used as a normative theory for how degrees of belief in a hypothesis might be revised in the light of new information.

Bayes' Theorem states that the odds of a hypothesis being correct in the light of new information (**posterior odds**) is the product of two elements: (1) the **prior odds** (the initial odds) of the hypothesis being correct *before* the information is observed multiplied by (2) the **likelihood ratio** – the ratio of the probabilities that, given the information, the hypothesis is correct ( $H$ ) or incorrect ( $\bar{H}$ ).

$$\frac{p(H|D)}{p(\bar{H}|D)} = \frac{p(D|H)}{p(D|\bar{H})} \times \frac{p(H)}{p(\bar{H})} \quad (H = \text{hypothesis}; D = \text{datum})$$

You should note that  $p(H|D)$  stands for the probability that the hypothesis is true, given that the datum or information  $D$  is true (read the vertical line as 'given'). Then, reading the formula from left to right, the three terms are:

- 1 The posterior odds ratio of  $H$  and  $\bar{H}$  (not  $H$ ) being true given  $D$ .
- 2 The likelihood ratio, representing the information value of  $D$  (the datum or information).
- 3 The prior odds of  $H$  and  $\bar{H}$  (not  $H$ ) being true before  $D$  is known.

If the probability of observing  $D$  when  $H$  is true is different from the probability of observing  $D$  when  $H$  is not true, then the information is diagnostic and the posterior odds will be different to the prior odds. The equation can be applied to any pair of competing hypotheses ( $A$  and  $B$ ) by replacing  $H$  and  $\bar{H}$  with  $A$  and  $B$ .

of drawing a second chip would now be applied to our new prior ratio to compute another posterior odds ratio. If it was yet another red chip belief in bag A would be  $\frac{0.7}{0.3} \times \frac{0.7}{0.3} = \frac{0.49}{0.09}$  – roughly 85%:15% – even more favourable for bag A. If the second chip was blue the information for each bag would be equivocal – our new posterior odds ratio would be  $\frac{0.3}{0.7} \times \frac{0.7}{0.3} = \frac{0.21}{0.21}$  – 50:50 odds that it is bag A or bag B.

A crucial aspect of the logic of these studies is that the experimenter is able to compare the subjective probabilities estimated by subjects with the objective probabilities calculated using Bayes' Theorem. All of the information required as inputs to Bayes' Theorem is explicit and unambiguous. Ironically though, this underplays the importance of the *subjectivity* probabilities. Because the experimenters assumed that they could objectively compute *the* correct answer, they reasoned that the subjective probabilities should be the same for all subjects faced with the same evidence. Calculating the objective probabilities and using them as a comparison then was absolutely necessary in order to be able to assess the accuracy

of people's judgements. However, it also indicates the artificiality of this kind of task, and is at the root of the difficulties that were to emerge with interpreting subjects' behaviour.

## 5.2 Does Bayes' Theorem describe human judgement?

The experiments conducted with the procedure discussed above produced ample evidence that human judgement under these conditions is not well described by Bayes' Theorem. Although subjects' revisions of their probability judgements were proportional to the values calculated from Bayes' Theorem, they did not revise their opinions sufficiently in the light of the evidence – a phenomenon that was labelled **conservatism**. The clear suggestion was that human judgement was poor, although there was some debate as to the precise reason for this. Perhaps it was due to a failure to understand the impact of the evidence or to an inability to combine the probability estimates according to Bayes' Theorem.

Aside from the theoretical interest in these possibilities, there were practical implications of this debate. If people are good at assessing probabilities but poor at combining them (as Edwards [1968] suggested) then perhaps they could be helped; a relatively simple remedy would be to design a support system that took the human assessments and combined them using Bayes' Theorem. However, if people were poor at assessing the component probabilities, then there wouldn't be much point in devising systems to help them combine these. 'Garbage in – garbage out' was a popular aphorism for summarizing this possibility.

However, before any firm conclusions were reached as to the cause of conservatism, the research exploring the phenomenon fizzled out. Two reasons for this can be identified. One cause (which we will consider in Sections 5.3 and 5.4) was the emergence of research into heuristics and biases and, in particular, the discovery of what Kahneman and Tversky (1973) called base-rate neglect. Before this development, however, growing disquiet was being voiced about the validity of book bag and poker chip experiments for assessing judgement.

Several studies had shown that quite subtle differences in the way that the tasks were presented to subjects resulted in considerable variability in the amount of conservatism. For example, the **Diagnosticity** of the data seemed an important variable. Diagnosticity of information means how much impact it has on opinion revision. Diagnosticity is indicated by the likelihood ratio. Imagine, instead of our two bags with a 70/30 split in the proportions of blue and red poker-chips, the bags contained 51 chips of one colour and 49 of the other. Clearly, two consecutive draws of a red chip would not be very diagnostic as to which of the two bags was being sampled. Phillips and Edwards' (1966) experiments showed that the more diagnostic the information, the more conservative was the subject. But when the information was very weakly diagnostic, as in this particular example, human probability revision, far from being conservative, was too extreme.

Another important factor was how the information was presented. Presenting information all at once or one bit at a time is irrelevant according to Bayes' Theorem but Peterson *et al.* (1965) found that presenting one item of information at a time, eliciting revisions after each item, produced less conservatism than giving the same information all in one go. Pitz *et al.* (1967) described this as an **inertia effect**: if an initial sequence of information favoured one of the hypotheses under evaluation,

subjects tended not to reduce their belief when confronted with later conflicting information.

DuCharme and Peterson (1968) investigated probability revision in a situation they considered nearer to real life than the standard paradigm. Most experiments, they complained, usually restricted information to one of two discrete possibilities (red or blue chip). In the real world, information leading to revision of opinion does not have discrete values but varies along a continuum. They gave their subjects the task of deciding which population was being sampled from – males or females – on the basis of the information given by randomly sampling heights from one of the populations. Using this task, DuCharme and Peterson found conservatism reduced to half the level found in the more artificial tasks. They concluded that this was due to their subjects' greater familiarity with the height distributions underlying their task.

Winkler and Murphy (1973) expressed further doubt concerning the validity of the conclusions from the book bag and poker chip paradigms. They argued that the standard task differed in several crucial aspects from the real world:

- 1 The pieces of evidence usually presented to subjects are conditionally independent. That is, knowing one piece of information does not change the likelihood of the other: producing one red chip from a bag, and then replacing it, does not affect the likelihood of drawing another red chip. However, in real world situations this assumption often does not make sense. For example, someone trying to discriminate hostile from friendly aircraft might spot an aircraft flying a non-standard route that fails to respond to radio signals. Flying off course and failing to respond are not independent – both could be caused by equipment failure. So, after observing one we should be less influenced by the other.

Winkler and Murphy argued that in many real-world situations lack of conditional independence of the information renders much of it redundant. In the standard tasks, subjects may have treated the information as if it was conditionally dependent and so one possible explanation for conservatism is that subjects are behaving much as they would do in familiar situations that involve redundant information sources.

- 2 In most experiments, the contents of the bags are fixed but in reality our hypotheses are not always constant; indeed, evidence may cause us to change the set of hypotheses under consideration.
- 3 In reality, information may be somewhat unreliable and therefore less diagnostic than the perfectly reliable colours of the poker-chips.
- 4 Typical experiments offer very diagnostic evidence – clearly favouring one hypothesis – whereas in reality evidence is very often weakly diagnostic. Again the result of generalizing from experience may be the appearance of conservatism. You will recall Phillips and Edwards' (1966) discovery that probability revision was too extreme with very weakly diagnostic evidence.

Winkler and Murphy concluded that 'conservatism may be an artifact caused by dissimilarities between the laboratory and the real world'.

## 5.3 Heuristics and biases

From the early 1970s Kahneman and Tversky provided a plethora of demonstrations of human judgemental error and linked these to the operation of a set of **mental heuristics** – mental rules of thumb – that they proposed the mind uses to simplify the process of judgement. These foibles, they argued, indicated that the underlying processes of judgement were not normative (e.g. did not compute probabilities using Bayes' Theorem) but instead used simpler rules that were easier for the brain to implement quickly.

The logic of their empirical research was to infer the characteristics of the mental processes underlying judgement by studying persistent biases – those not due to inattention or fatigue. The idea, spelled out in Kahneman *et al.* (1982), is that, due to limited mental processing capacity, strategies of simplification are required to reduce the complexity of judgement tasks and render them tractable by the kind of mind that people have. Accordingly, the principal reason for interest in judgmental biases was not merely that subjects made errors but that the errors revealed how people made use of relatively simple but error-prone heuristics for making judgements.

### 5.3.1 The ‘representativeness’ heuristic

The **representativeness** heuristic is used to determine how likely it is that an event is a member of a category by considering how similar or typical the event is to the category (remember the similarity-based approach to categorization discussed in Chapter 5?). For example, people may judge the likelihood that a given individual is a librarian by the extent to which the individual resembles a ‘typical’ librarian. This may seem a reasonable strategy but it neglects consideration of the relative prevalence of librarians in society as a whole: the so-called **base rate**. We have seen that Bayes' Theorem prescribes that prior likelihood is an important component when assessing the impact of new information. So, when given information about an individual, the chances that he or she is a member of a profession will still be influenced by the prior likelihood – or base rate – for that profession. Knowing that someone regularly works in the British Library might increase your belief that they are a famous writer, but it is still more likely that she or he is a librarian because there are more of them than famous writers. Tversky and Kahneman found that when base rates of different categories varied, judgements of the occupations of described people were correspondingly biased – due to base-rate neglect. People using the representativeness heuristic for forecasting were employing a form of stereotyping in which similarity dominated other cues as a basis for judgement.

In Kahneman and Tversky's (1973) experiments demonstrating neglect of base rates, subjects were found to ignore information concerning the prior probabilities of the hypotheses – the polar opposite of conservatism. For example, in one study subjects were presented with this brief personal description of an individual called Jack.

Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing and mathematical puzzles.

Half the subjects were told that the description had been drawn from a sample of seventy engineers and thirty lawyers while the other half were told that the description was drawn from a sample of thirty engineers and seventy lawyers. So, the base rate, or prevalence of engineers for the two groups was 70 per cent and 30 per cent respectively. However, when asked to estimate the probability that Jack was an engineer, the mean estimates of the two groups were only very slightly different (50 per cent vs 55 per cent). On the basis of such results, Kahneman and Tversky concluded that prior probabilities are largely ignored when individuating information is made available.

Kahneman and Tversky then gave a description designed to be totally uninformative about the profession of the individual:

Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field.  
He is well liked by his colleagues.

When contemplating this description, subjects given markedly different base rates produced identical median estimates of 50 per cent. Kahneman and Tversky concluded that base rates were properly utilized when no specific information was given, but that base rates were neglected when even worthless information was provided (as in this example).

Tversky and Kahneman (1983) also invoked judgement by representativeness to explain the **conjunction fallacy** whereby a conjunction of two events is judged to be more likely than one of those events alone. The fallacy violates a simple principle of probability: the probability of a conjunction A and B can never exceed either the probability of A or the probability of B. Nevertheless, subjects who read a description of a woman called Linda who had a history of interest in liberal causes thought it more likely that she was a feminist bank clerk (i.e. a conjunction – Linda is a feminist *and* a bank clerk) than just a bank clerk, thereby violating the conjunction rule. Of course, though all feminist bank clerks are bank clerks, feminist bank clerks are more *representative* of people interested in liberal causes than bank clerks in general. So, while valid probabilities respect the conjunction rule, judgements of representativeness may not.

### 5.3.2 The 'availability' heuristic

The **availability** heuristic is invoked when people estimate likelihood or relative frequency by the ease with which instances can be brought to mind. Instances of frequent events are typically easier to recall than instances of less frequent events so availability will often be a valid cue for estimates of likelihood. However, availability is affected by factors other than likelihood. For example, recent events and emotionally salient events are easier to recollect. It is a common experience that the perceived riskiness of air travel arises in the immediate wake of an air disaster.

Judgements made on the basis of availability then are vulnerable to bias whenever availability and likelihood are uncorrelated.

### 5.3.3 The ‘anchor and adjust’ heuristic

The **anchor and adjust** heuristic is used when people make estimates by starting from an initial value and then adjust it to arrive at their final estimate. The claim is that adjustment is typically insufficient. For instance, one experimental task required subjects to estimate various quantities stated in percentages (e.g. the percentage of African countries in the UN). Subjects communicated their answers by using a spinner wheel showing numbers between 0 and 100. For each question, the wheel was spun and then subjects were first asked whether the true answer was above or below this arbitrary value. They then gave their estimate of the actual value. Perversely, people’s estimates were found to correlate with the initial (entirely random) starting point (cf. Wilson *et al.*, 1996).

## 5.4 Evaluating the heuristics and biases account

The heuristics and biases research provided a methodology, a vivid explanatory framework and a strong suggestion that judgement is not as good as it might be. Kahneman and Tversky (1982) made clear that the main goal of their research was to understand the processes that produce both valid and invalid judgements. However, it soon became apparent that ‘although errors of judgement are but a method by which some cognitive processes are studied, the method has become a significant part of the message’ (Kahneman and Tversky, 1982, p.494). So how should we regard human judgement?

There has been an enormous amount of discussion of Tversky and Kahneman’s findings and claims. Researchers in the heuristics and biases tradition have generated shock and astonishment that people seem so bad at judging probability despite the fact that we all live in an uncertain world. Not surprisingly, these claims have been challenged. Some question whether the demonstrations of biases in judgement apply to experts operating in their domain of expertise or merely to student samples. Another argument is that the experimental tasks set to subjects provide a misleading perspective of their competence. A third argument is that the standards for the assessment of judgement are inappropriate.

Consideration of a prominent critique of Tversky and Kahneman’s argument is given below.

### 5.4.1 Representativeness and base-rate neglect

Following Tversky and Kahneman’s original demonstration of base-rate neglect, research established that base rates might be attended to more (though usually not sufficiently) if they were perceived as relevant (Bar-Hillel, 1980), had a causal role (Kahneman and Tversky 1982), or were ‘vivid’ rather than ‘pallid’ (Nisbett and Ross, 1980). However, Gigerenzer *et al.* (1988) argued that the variations in base-rate neglect have nothing to do with any of these factors *per se*, but arise because different problems may to varying degrees encourage the subject to represent the problem as a Bayesian revision problem. Just because the experimenter assumes that she has defined a probability problem does not imply that the subject will see it in the same way. In particular, subjects may have reasons not to take the base rate asserted

by the experimenter as their subjective prior probability. In Kahneman and Tversky's original experiments the descriptions were not actually randomly sampled (as the subjects were told) but especially selected to be 'representative' of the professions. To the extent that subjects suspected this was the case then they would be entitled to ignore the offered base rate.

Gigerenzer *et al.* (1988) let their subjects experience the sampling themselves. Their subjects examined ten pieces of paper each marked lawyer or engineer in proportion to the base rates. Subjects then drew one of the pieces of paper from an urn and unfolded it so they could read a description of an individual without being able to see the mark defining it as being of a lawyer or engineer. In these circumstances, subjects used the base rates in a proper fashion – base-rate neglect 'disappeared'. However in a replication where base rates were asserted, rather than sampled, Kahneman and Tversky's base-rate neglect was replicated.

Kahneman and Tversky (1996) have argued that a fair summary of the research would be that explicitly presented base rates are generally under-weighted but not ignored. They also pointed out that, in Gigerenzer *et al.*'s (1988) experiment, subjects who sampled the information themselves still produced judgements that deviated from the Bayesian solution in the direction predicted by representativeness. Evidently then representativeness is useful for predicting judgements. However, to the extent that base rates are not entirely ignored (Koehler, 1995), the heuristic rationale for representativeness is limited. You will recall that the original explanation for base-rate neglect was the operation of a simple heuristic that reduced the need for integration of information. If judgements in these experiments reflect use of base rates – albeit to a limited extent – it is hard to account for findings by the operation of a simplifying representativeness heuristic.

#### 5.4.2 Frequency and the conjunction fallacy

Tversky and Kahneman (1983) reported evidence that violations of the conjunction rule largely disappeared when subjects were requested to assess the relative **frequency** of events rather than the probability of a single event. Thus, instead of being asked about the likelihood for a particular individual, subjects were requested to assess how many people in a survey of 100 adult males had had heart attacks and then were asked to assess the number who were both over 55 years old *and* had had heart attacks. Only 25 per cent of subjects violated the conjunction rule by giving higher values to the latter than to the former. When asked about likelihoods for single events, however, it is typically the vast majority of subjects who violate the rule. This difference in performance between frequency and single-event versions of the conjunction problem has been replicated several times since (cf. Gigerenzer, 1994).

Gigerenzer (1994) has suggested that people are naturally adapted to reasoning with information in the form of frequencies and that because of this the conjunction fallacy 'disappears' if reasoning is based on frequencies. This suggests that the difficulties that people experience in solving probability problems can be reduced if the problems require subjects to assess relative frequency for a class of events rather than the probability of a single event. Thus, it is possible that if judgements were elicited with frequency formats there would be no biases. Kahneman and Tversky (1996) disagree and argue that the frequency format serves to provide subjects with a powerful cue to the relation of inclusion between sets that are explicitly compared, or

evaluated in immediate succession. When the structure of the conjunction is made more apparent, then subjects who appreciate the constraint supplied by the rule will be less likely to violate it. According to their account, salient cues to set inclusion – not the frequency information *per se* – prompted subjects to adjust their judgement.

To test this explanation, Kahneman and Tversky (1996) reported a new variation of the conjunction problem experiment where subjects made judgements of frequencies but the cues to set inclusion were removed. They presented subjects with the description of Linda and then asked their subjects to suppose that there were 1,000 women who fit the description. They then asked one group of subjects to estimate how many of them would be bank tellers; a second, independent group of subjects were asked how many were bank tellers and active feminists; a third group made evaluations for both categories. As predicted, those subjects who evaluated both categories mostly conformed to the conjunction rule. However, in a between-groups comparison of the other two groups, the estimates for 'bank tellers and active feminists' were found to be significantly higher than the estimates for bank tellers. Kahneman and Tversky argue that these results show that subjects use the representativeness heuristic to generate their judgements and then edit their responses to respect class inclusion where they detect cues to that relation. Thus, they concluded that the key variable controlling adherence to the conjunction rule is not the relative frequency format *per se* but the opportunity to detect the relation of class inclusion.

Other authors have investigated the impact of frequency information (Evans *et al.*, 2000; Girotto and Gonzales, 2002) and concluded that it is not the frequency information *per se* but the perceived relations between the entities that is affected by different versions of the problem, though this is rejected by Hoffrage *et al.* (2002). We need to understand more of the reasons underlying the limiting conditions of cognitive biases – how it is that seemingly inconsequential changes in the format of information can so radically alter the quality of judgement. Biases that can apparently be cured so simply cannot plausibly be held to reveal fundamental and immutable characteristics of judgement processes. We shall now consider the history of one well-known cognitive bias: overconfidence.

## 5.5 Overconfidence

In the 1970s and 1980s a considerable amount of evidence was marshalled for the view that people suffer from an overconfidence bias. Typical laboratory studies of calibration ask subjects to answer a question such as:



Subjects are required to indicate the answer that they think is correct and then state how confident they are on a probability scale ranging from 50 per cent to 100 per cent (the minimum is 50 per cent since one of the answers is always correct and 50 per cent is the probability of guessing correctly). To be well calibrated, an assessed probability should correspond with the number of correct judgements over a number of assessments. For example, if you assign a probability of 70 per cent to each of ten predictions then you should get seven of those predictions correct. Typically, however, people tend to give **overconfident** responses – their average confidence is

higher than their proportion of correct answers. McClelland and Bolger (1994) and Harvey (1997) give comprehensive reviews of this aspect of probabilistic judgement.

Overconfidence has been recorded in the judgements of experts. For example, Christensen-Szalanski and Bushyhead (1981) explored the validity of the probabilities given by physicians to diagnoses of pneumonia. They found that the probabilities were poorly calibrated and very overconfident; the proportion of patients who turned out to have pneumonia was far less than the probability statements implied. Wagenaar and Keren (1986) found overconfidence in lawyers' predictions of the outcome of court trials in which they represented one side. As they point out, it is inconceivable that the lawyers do not pay attention to the outcomes of trials in which they have participated, so why don't they learn to make well-calibrated judgements?

Could the circumstances in which some experts operate impede the proper monitoring of feedback necessary for the development of well-calibrated judgements? A consideration of the reports of well-calibrated experts supports this notion; they all appear to be cases where some explicit unambiguous quantification of uncertainty is routinely made and the outcome feedback is prompt and unambiguous. Doctors and lawyers don't routinely quantify their uncertainty, and may have to wait months to discover the outcomes of their judgements, the truth of which may never be revealed.

The most commonly cited example of well-calibrated judgements is weather forecasters' estimates of the likelihood of rainfall (Murphy and Winkler, 1984) but there are others. Keren (1987) found highly experienced tournament bridge players (but not experienced non-tournament players) made well-calibrated forecasts of the likelihood that their bids would be made, and Phillips (1987) reports well-calibrated forecasts of horse races by bookmakers. In each of these three cases, the judgements made by the experts are precise numerical statements and the outcome feedback is unambiguous and received promptly and so can be easily compared with the initial forecast. Under these circumstances, experts are unlikely to be insensitive to the experience of being surprised; there is very little scope for neglecting, or denying, any mismatch between forecast and outcome.

Following the ideas of Brunswik (1943, 1955) – that cognition is well adapted to people's natural environments – some judgement researchers have argued that overconfidence is an artifact of artificial experimental tasks and the non-representative sampling of stimulus materials. Gigerenzer *et al.* (1991) and Juslin (1994) claim that overconfidence is observed because the typical general knowledge quizzes used in most experiments contain a disproportionate number of misleading items. For example, most people judge wrongly that Rome is south of New York. These authors found that when knowledge items are randomly sampled the overconfidence phenomenon disappears. Gigerenzer *et al.* (1991) presented their subjects with randomly selected pairs of German cities. When asked to select the biggest and indicate their confidence, overconfidence was not observed.

Erev *et al.* (1994) spotted another misleading source of evidence of overconfidence. They explained that overconfidence might, to some degree, reflect an underlying random component of judgement. When any two variables are not perfectly correlated – and confidence and accuracy aren't *perfectly* correlated – there

will be a regression effect. For example, the heights of fathers and their sons are positively – but not perfectly – correlated. Consequently, a sample of the (adult) sons of extremely tall fathers will, on average, be shorter than their fathers *and at the same time* a sample of the fathers of extremely tall (adult) sons will, on average, be shorter than their sons. Thus, depending on how you sampled – either looking at very tall sons or very tall fathers – could lead you to two opposite conclusions about the direction of a (non-existent) difference between the populations.

So could it really be that all the evidence for overconfidence is merely an illusion created by inappropriate sampling of test items and regression effects? Budescu *et al.* (1997) attempted to measure and control for the regression effects caused by random variation in judgements by presenting the same items (random pairs of large American cities) on several occasions to their subjects. They found that the vast majority of the individuals in their study (87 per cent) were biased towards overconfidence even after the effects of random error in their judgements had been taken into account. As they also used a representative sample of items, both the artefactual sources of overconfidence should have been eliminated.

Juslin *et al.* (2000) report a meta-analysis comparing 35 studies, where the items for judgement were randomly selected from a defined domain, with 95 studies where items were selected non-randomly by experimenters. While overconfidence was evident for selected items, it was close to zero for randomly sampled items, which suggests that overconfidence is not simply a ubiquitous cognitive bias. This analysis suggests that the appearance of overconfidence may be an illusion – not one experienced by experimental subjects, but one inadvertently created and suffered by researchers, and so not a cognitive bias in their respondents.

## Summary of Section 5

---

- There have been many demonstrations of human judgmental error, linking errors to the operation of various heuristics and biases.
- There is evidence that many of these reported biases disappear when the wording of the problems is changed. Whether this reflects sensitivity to particular cues or adaptivity to frequency information formats is the subject of debate.

## 6 Fast and frugal theories of decision making

Another approach to judgement and choice that has recently emerged tests the efficacy of heuristics on information occurring ‘in the wild’ rather than on specially contrived laboratory problems. Gigerenzer and Goldstein (1996) comparatively evaluated the performance of a set of different decision strategies. Instead of focusing on violations of normative rules, they produced a measure of the efficacy of simple mental strategies for judgement by measuring the number of correct inferences that different strategies made. The class of simple models that Gigerenzer

and Goldstein tested were what they called **fast and frugal** heuristics: ‘frugal’ because these heuristics used just one piece of information in order to make decisions; ‘fast’ because they didn’t attempt any sort of integration of different bits of information prescribed by such normative procedures as SEU or Bayes’ Theorem. By the standards of classical rationality enshrined in normative rules, the mental strategies that Gigerenzer and Goldstein considered look very primitive. Indeed, they were quite explicit about the fact that the simple heuristics that they tested violate basic axioms such as transitivity. Nevertheless, the proof of the pudding is in the eating; as we have seen, people and bees violate transitivity and yet manage to get by.

The inspiration for this exercise was Simon’s (1956) idea of **bounded rationality**. Simon emphasized that, due to its limited capacity, human information processing would be obliged to use **satisficing** methods for problem solving (satisficing is an old Northumbrian word meaning to satisfy). Simon used it to describe decision procedures that, while not optimal, reflect the constraints supplied by human information-processing capacity and the opportunities provided by the structure of the environment. Most research on human judgement has focused on the non-optimal nature of simple human information-processing strategies – the importance of the environment structure in determining performance has been overlooked. Nevertheless, we have seen how evidence for both conservatism and overconfidence was undermined by considering how mental strategies might exploit the way information is structured in the natural environment. But could judgement strategies that violate normative rules and utilize just one piece of information possibly be of effective service?

In their study, Gigerenzer and Goldstein used the properties of a set of German cities as information on which to base decisions as to which city was the biggest. Commonly known correlates of city size such as whether it is the state capital, has a university, a football club in the top division or an inter-city rail station were the cues that the heuristics could use. One heuristic that Gigerenzer and Goldstein tested they called ‘Take the Best’ – so called as it simply worked through the cues in order of their predictive validity until one was found that discriminated between two cities and then responded accordingly. Thus, if the two cities under consideration could not be discriminated on the basis of the most diagnostic cue (e.g. whether it is the state capital or not) the search through memory continues. The search for discriminatory cue values proceeds in order of their relative diagnosticity until a cue is found that discriminates the two cities (e.g. one has an inter-city rail station and the other does not), whereupon information retrieval is stopped and the judgement made according to this single cue.

Gigerenzer and Goldstein compared simple heuristics such as ‘Take the Best’ with other decision rules that integrate multiple bits of information (such as multiple regression). They modelled the effect of limited knowledge by simulating six classes of subjects who knew varying proportions of the cue values associated with the cities. Surprisingly, they found that ‘Take the Best’ did as well as any of the other algorithms and considerably better than some. As it only uses one piece of information it would be much faster than any process that retrieves multiple bits of information and attempts integration of the information. The result is important for

demonstrating that, although adherence to normative rules may be sufficient for good judgement, it is not necessary.

Further demonstrations of the efficacy of fast and frugal heuristics have studied binary decisions in a wide range of types of knowledge environment (e.g. which professor has the highest salary? Which US city has more homeless people?). These studies extended the application beyond choice to value estimation, categorization and memory (Gigerenzer *et al.*, 1999). Goldstein and Gigerenzer (2002) asked American and German students which is bigger: San Antonio or San Diego? While 62 per cent of the Americans correctly named San Diego, 100 per cent of the German students were correct. The Germans were applying a **recognition heuristic** – if you recognize one and not the other, pick the city you have heard of. As you usually hear about the bigger cities of foreign countries before the smaller ones this will be a pretty good cue. Because the Americans had heard of both cities they couldn't apply this cue and had to rely on other, apparently less valid, cues. In the same way, for city-size decisions, American students were slightly more accurate about German cities than American cities. Ignorance can even sometimes be helpful because simple mental heuristics *can* exploit the structure of information in the environment to make good inferences. As a consequence of such results, we might question the present pre-eminent status of normative rules for defining rationality and for serving as a benchmark for assessing human judgement.

## Summary of Section 6

---

- Human decision making may employ fast and frugal heuristics – simple rules that yield quick decisions yet which can be highly accurate in certain natural environments.

## 7 Conclusion

The idea that people don't decide as they should was appreciated by psychologists very early on. In a seminal paper, which effectively introduced the study of decision making to psychology, Edwards (1954, p.382) wrote: 'It is easy for a psychologist to point out that an economic man ... is very unlike a real man.' Yet for economists this disparity is less clear. As Lopes (a psychologist) put it: 'Economics considers itself a normative science, the very term an oxymoron of ought and is' (1994, p.222). Psychologists and economists think rather differently about the behavioural research exploring decision making (cf. Hertwig and Ortmann, 2001; Lopes, 1994). To psychologists, it is evident that people cannot conceivably represent all the relevant information that normative models require for judgement and decision: 'Who could design a brain that could perform the way this model mandates? Every single one of us would have to know and understand everything completely, and at once' (Daniel Kahneman quoted by Bernstein, 1996).

While it may be a tad optimistic to presume that the social sciences are on the verge of reconciliation and consensus on this subject, psychology has, since the

1950s, made enormous progress in establishing that actual human decision making cannot be satisfactorily characterized in the idealized way mathematics and economics have assumed. Moreover, alternative descriptive theories that account for the discrepancies are emerging. Perhaps the best recent piece of evidence for that claim is that a psychologist – Daniel Kahneman – shared the 2002 Nobel Prize for Economics ‘for having integrated insights from psychological research into economic science, especially concerning human judgement and decision making under uncertainty’ (Nobel citation, 2002).

Does the rejection of normative theory as a model for human judgement and choice imply that judgement and choice must be poor or even ‘irrational’? No. Although the evidence that people do not perform ideally is clear, any reasonable standards of rationality must surely accept that the computational requirements of normative models are beyond the capacity of a human brain: nevertheless, such bounded rationality (cf. Simon, 1956) does not imply irrationality.

In my (admittedly fallible) judgement the issues surrounding the nature and evaluation of human judgement and decision making are profound and will not be resolved easily or in the near future. To make further progress, we need studies that do more than merely knock down the straw man defined by normative models. Among the many questions that arise, two broad issues can be framed: first, how is it that we are as competent as we evidently are? Second, what can we do about how incompetent we evidently are? Quite how it is that people perform as effectively as they do by applying non-normative mental strategies to the limited information that they can process – and how we might learn to improve our decision making – remain to be explored and explained.

## Further reading

- Gigerenzer, G. and Selten, R. (eds) (2001) *Bounded Rationality: The Adaptive Toolbox*, Cambridge, MA., MIT Press.
- Kahneman, D. and Tversky, A. (eds) (2000) *Choices, Values and Frames*, Cambridge, Cambridge University Press.
- Koehler, D. and Harvey, N. (eds) (2004) *Blackwell Handbook of Judgment and Decision Making*, Oxford, Blackwell Publishing.

## References

- Allais, M. (1953) ‘Le comportement de l’homme rationnel devant le risque, critique des posulats et axioms de l’école Américaine’, *Econometrica*, vol.21, pp.503–46.
- Allais, M. (1979) ‘The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school’, in Allais, M. and Hagen, O. (eds) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.
- Allais, M. and Hagen, O. (eds) (1979) *op. cit.*, Dordrecht, Reidel.
- Bar-Hillel, M. (1980) ‘The base-rate fallacy in probability judgements’, *Acta Psychologica*, vol.44, pp.211–33.

- Bernstein, P.L. (1996) *Against the Gods: The Remarkable Story of Risk*, New York, Wiley.
- Brunswik, E. (1943) ‘Organismic achievement and environmental probability’, *Psychological Review*, vol.50, pp.255–72.
- Brunswik, E. (1955) ‘Representative design and probabilistic theory in a functional psychology’, *Psychological Review*, vol.62, pp.193–217.
- Budescu, D., Wallsten, T.S. and Au, W.T. (1997) ‘On the importance of random error in the study of probabilistic judgement: Part II: Applying the stochastic judgement model to detect systematic trends’, *Journal of Behavioral Decision Making*, vol.10, pp.173–88.
- Christensen-Szalanski, J.J.J. and Bushyhead, J.B. (1981) ‘Physicians’ use of probabilistic information in a real clinical setting’, *Journal of Experimental Psychology, Human Perception and Performance*, vol.7, pp.928–35.
- Cohen, L.J. (1981) ‘Can human irrationality be experimentally demonstrated?’, *The Behavioral and Brain Sciences*, vol.4, pp.317–70.
- DuCharme, W.M. and Peterson, C.R. (1968) ‘Intuitive inference about normally distributed populations’, *Journal of Experimental Psychology*, vol.78, pp.269–75.
- Edwards, W. (1954) ‘The theory of decision making’, *Psychological Bulletin*, vol.41, pp.380–417.
- Edwards, W. (1955) ‘The prediction of decisions among bets’, *Journal of Experimental Psychology*, vol.50, pp.201–14.
- Edwards, W. (1968) ‘Conservatism in human information processing’ in Kleinmuntz, B. (ed.) *Formal Representation of Human Judgment*, New York, Wiley.
- Edwards, W. (1992) ‘Toward the demise of economic man and woman: bottom lines from Santa Cruz’ in Edwards, W. (ed.) *Utility Theories: Measurements and Applications*, Dordrecht, Kluwer.
- Erev, I., Wallsten, T.S. and Budescu, D.V. (1994) ‘Simultaneous over-and underconfidence: the role of error in judgement processes’, *Psychological Review*, vol.101, pp.519–28.
- Evans, J. St. B.T., Handley, S.J., Perham, N., Over, D.E. and Thompson, V.A. (2000) ‘Frequency versus probability formats in statistical word problems’, *Cognition*, vol.77, pp.197–213.
- Gigerenzer, G. (1994) ‘Why the distinction between single event probabilities and frequencies is important for psychology and vice-versa’ in Wright, G. and Ayton, P. (eds) *Subjective Probability*, Chichester, Wiley.
- Gigerenzer, G., Hell, W. and Blank, H. (1988) ‘Presentation and content: the use of base rates as a continuous variable’, *Journal of Experimental Psychology: Human Perception and Performance*, vol.14, pp.513–25.
- Gigerenzer, G., Hoffrage, U. and Kleinbölting, H. (1991) ‘Probabilistic mental models: A Brunswikian theory of confidence’, *Psychological Review*, vol.98, pp.506–28.
- Gigerenzer, G. and Goldstein, D.G. (1996) ‘Reasoning the fast and frugal way: Models of bounded rationality’, *Psychological Review*, vol.103, pp.650–69.

- Gigerenzer, G., Todd, P.M. and The ABC Research Group (1999) *Simple Heuristics that Make us Smart*, Oxford, Oxford University Press.
- Girotto, V. and Gonzales, M. (2002) 'Chances and frequencies in probabilistic reasoning: rejoinder to Hoffrage, Gigerenzer, Krauss, and Martignon', *Cognition*, vol.84, pp.353–9.
- Goldstein, D.G. and Gigerenzer, G. (2002) 'Models of ecological rationality: the recognition heuristic', *Psychological Review*, vol.109, pp.75–90.
- Grether, D.M. and Plott, C.R. (1979) 'Economic theory of choice and the preference reversal phenomenon', *The American Economic Review*, vol.69, pp.623–38.
- Harvey, N. (1997) 'Confidence in judgement', *Trends in Cognitive Sciences*, 1, pp.78–82.
- Hertwig, R. and Ortmann, A. (2001) 'Experimental practices in economics: a methodological challenge for psychologists?' *Behavioral and Brain Sciences*, vol.24, p.383.
- Hoffrage, U., Gigerenzer, G., Krauss, S. and Martignon, L. (2002) 'Representation facilitates reasoning: what natural frequencies are and what they are not', *Cognition*, vol.84, pp.343–52.
- Hsee, C.K. (1998) 'Less is better; When low-value options are valued more highly than high-value options', *Journal of Behavioral Decision Making*, vol.11, pp.107–21.
- Juslin, P. (1994) 'The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items', *Organizational Behavior and Human Decision Processes*, vol.57, pp.226–46.
- Juslin, P., Winman, A. and Olsson, H. (2000) 'Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect', *Psychological Review*, vol.107, pp.384–96.
- Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.
- Kahneman, D. and Tversky, A. (1972) 'Subjective probability: a judgement of representativeness', *Cognitive Psychology*, vol.3, pp.430–54.
- Kahneman, D. and Tversky, A. (1973) 'On the psychology of prediction', *Psychological Review*, vol.80, pp.237–51.
- Kahneman, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decision making under risk', *Econometrica*, vol.47, pp.263–91.
- Kahneman, D. and Tversky, A. (1982) 'On the study of statistical intuitions' in Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *op. cit.*
- Kahneman, D. and Tversky, A. (1996) 'On the reality of cognitive illusions: a reply to Gigerenzer's critique', *Psychological Review*, vol.103, pp.582–91.
- Keller, L.R. (1985) 'The effects of problem representation on the sure-thing and substitution principles', *Management Science*, vol.31, pp.738–51.
- Keren, G.B. (1987) 'Facing uncertainty in the game of bridge: a calibration study', *Organizational Behavior and Human Decision Processes*, vol.39, pp.98–114.
- Koehler, J.J. (1995) 'The base-rate fallacy reconsidered – descriptive, normative, and methodological challenges', *Behavioral and Brain Sciences*, vol.19, pp.1–55.

- Lichtenstein, S. and Slovic, P. (1973) ‘Response-induced reversals of preference in gambling: an extended replication in Las Vegas’, *Journal of Experimental Psychology*, vol.101, pp.16–20.
- Lichtenstein, S., Slovic, P. and Zink, D. (1969) ‘Effect of instruction in expected value on optimality of gambling decisions’, *Journal of Experimental Psychology*, vol.79, pp.236–40.
- Lopes, L.L. (1994) ‘Psychology and economics: Perspectives on risk, cooperation, and the marketplace’, *Annual Review of Psychology*, vol.45, pp.197–227.
- McClelland, A.G.R. and Bolger, F. ‘The calibration of subjective probabilities: theories and models 1980–1994’ in Wright, G. and Ayton, P. (eds) *Subjective Probability*, New York, Wiley.
- Murphy, A.H. and Winkler, R.L. (1984) ‘Probability forecasting in meteorology’, *Journal of the American Statistical Association*, vol.79, pp.489–500.
- Nisbett, R. and Ross, L. (1980) *Human Inference: Strategies and Shortcomings*, Englewood Cliffs, NJ, Prentice Hall.
- Nobel Foundation (2004) *The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel*, <http://www.nobel.se/economics/laureates/2002/index.html> (accessed 27 July 2004).
- Peterson, C.R., Schneider, R.J. and Miller, A.J. (1965) ‘Sample size and the revision of subjective probability’, *Journal of Experimental Psychology*, vol.69, pp.522–7.
- Phillips L.D. (1984) ‘A theory of requisite decision-models’, *Acta Psychologica*, vol.56, pp.29–48.
- Phillips, L.D. (1987) ‘On the adequacy of judgmental probability forecasts’ in Wright, G. and Ayton, P. (eds) *Judgemental Forecasting*, Chichester, Wiley.
- Phillips, L.D. (1989) ‘Decision analysis in the 1990s’ in Shahini, A. and Stainton, R. (eds) *Tutorial Papers in Operational Research 1989*, Birmingham, Operational Research Society.
- Phillips, L.D. and Edwards, W. (1966) ‘Conservatism in simple probability inference tasks’, *Journal of Experimental Psychology*, vol.72, pp.346–57.
- Pitz, G.F., Downing, L. and Rheinold, H. (1967) ‘Sequential effects in the revision of subjective probabilities’, *Canadian Journal of Psychology*, vol.21, pp.381–93.
- Raiffa, H. (1968) *Decision Analysis*, Reading, MA, Addison Wesley.
- Samuelson, P.A. (1963) ‘Risk and uncertainty: a fallacy of large numbers’, *Scientia*, vol.98, pp.108–13.
- Savage, L. (1954) *The Foundations of Statistics*, New York, Wiley.
- Schlaifer, R. (1969) *Analysis of Decisions under Uncertainty*, New York, McGraw Hill.
- Shafir, E. (1993) ‘Choosing versus rejecting: why some options are both better and worse than others’, *Memory and Cognition*, vol.21, pp.546–56.
- Shafir, E., Simonson, I. and Tversky, A. (1993) ‘Reason-based choice’, *Cognition*, vol.49, pp.11–36.
- Shafir, S. (1994) ‘Intransitivity of preferences in honeybees: support for “comparative” evaluation of foraging options’, *Animal Behaviour*, vol.48, pp.55–67.

- Simon, H.A. (1956) 'Rational choice and the structure of the environment', *Psychological Review*, vol.63, pp.129–38.
- Slovic, P. (1975) 'Choice between equally valued alternatives', *Journal of Experimental Psychology: Human Perception and Performance*, vol.1, pp.280–7.
- Slovic, P. (1995) 'The construction of preference', *American Psychologist*, vol.50, pp.364–71.
- Slovic, P., Fischhoff, B. and Lichtenstein, S. (1977) 'Behavioral decision theory', *Annual Review of Psychology*, vol.28, pp.1–39.
- Slovic, P. and Lichtenstein, S. (1968) 'Relative importance of probabilities and payoffs in risk taking', *Journal of Experimental Psychology Monograph*, vol.78, pp.1–18.
- Slovic, P. and Tversky, A. (1974) 'Who accepts Savage's axiom?', *Behavioral Science*, vol.19, pp.368–73.
- Tversky, A. (1969) 'Intransitivity of preferences', *Psychological Review*, vol.76, pp.31–48.
- Tversky, A. and Kahneman, D. (1974) 'Judgement under uncertainty: heuristics and biases', *Science*, vol.185, pp.1124–31.
- Tversky, A. and Kahneman, D. (1981) 'The framing of decisions and the psychology of choice', *Science*, vol.211, pp.453–8.
- Tversky, A. and Kahneman, D. (1983) 'Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement', *Psychological Review*, vol.90, pp.293–315.
- Tversky, A. and Kahneman, D. (1986) 'Rational choice and the framing of decisions', *Journal of Business*, vol.59, S251–S278.
- Tversky, A., Sattath, S. and Slovic, P. (1988) 'Contingent weighting in judgement and choice', *Psychological Review*, vol.95, pp.371–84.
- Tversky, A., Slovic, P. and Kahneman, D. (1990) 'The causes of preference reversal', *American Economic Review*, vol.80, pp.204–17.
- von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.
- von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioral Research*, Cambridge, Cambridge University Press.
- Wagenaar, W.A. and Keren, G.B. (1986) 'Does the expert know? The reliability of predictions and confidence ratings of experts' in Hollnagel, E., Mancini, G. and Woods, D.D. (eds) *Intelligent Decision Support in Process Environments*, Berlin, Springer-Verlag.
- Watson, S.R. and Buede, D.M. (1987) *Decision Synthesis: The Principles and Practice of Decision Analysis*, Cambridge, Cambridge University Press.
- Wilson, T.D., Houston, C.E., Etling, K.M. and Brekke, N. (1996) 'A new look at anchoring effects: basic anchoring and its antecedents', *Journal of Experimental Psychology: General*, vol.125, pp.387–402.
- Winkler, R.L. and Murphy, A.M. (1973) 'Experiments in the laboratory and the real world', *Organizational Behavior and Human Performance*, vol.20, pp.252–70.

Mike Oaksford

---

### 1 Introduction

Suppose a friend tells you that

1     *If John finds out, then he will be furious.*

and you discover later that

2     *John found out.*

If you then conclude that

3     *John was furious.*

you will have engaged in **reasoning** – that is, you will have inferred a **conclusion** (3) from some initial information or **premises** (1) and (2).

Reasoning has been studied since the time of the ancient Greeks. Aristotle suggested that reasoning is one of the abilities that marks us off from the other animals, implying that only humans are able to reason and only humans have minds that are capable of rational thought.

#### 1.1 Reasoning and logic

Aristotle also developed the first system of **logic**: he produced a set of rules by which to judge whether certain passages of reasoning, like the one above, were valid; that is, for telling whether the conclusion (3) really does follow from the premises (1) and (2). This is one of the sources of a strong line adopted by some researchers in this area: reasoning is the process of applying logical laws.

This strong line also emerges in the foundations of modern logic – for instance, in Boole's *The Laws of Thought* (1854). In that book, Boole described a set of rules that determine how we can draw inferences from statements like *if ... then* (as in Example 1). The book's title clearly reveals that the author's intention was to describe the laws that govern human reasoning.

More recently, Piaget placed the ability to reason according to logical rules at the pinnacle of his **stage theory** of cognitive development, that is, at the formal operational stage (Inhelder and Piaget, 1958). One of our main concerns in this chapter will be to determine whether logic provides a good model of human reasoning. This does not mean that we should expect people to reason perfectly logically. Although we may all be *capable* of reasoning logically, perfect performance may, for example, take up too much memory. In which case, we might expect some errors to emerge and, as we will see, such errors have been observed.

Another theme that will emerge is whether logic is appropriate to describe real human reasoning at all. The full title of Boole's (1854) book was *An Investigation of the Laws of Thought On Which Are Founded the Mathematical Theories of Logic and Probabilities*. Almost one half of the book was devoted to probability theory, which Boole thought may provide a better theory of everyday reasoning. As we will see, some researchers do not view people's reasoning behaviour as *error prone but logical*. Rather, they view it as relatively *error free but probabilistic*. This difference emerges as a result of assigning differing meanings to the important words that figure in a passage of reasoning. In this chapter, we will concentrate heavily on the construction *if ... then*. According to some researchers, *if ... then* should be interpreted logically, in terms of the conditions under which it is true or false. For example, (1) is false if John does not get furious when he finds out; otherwise, it is true. Other researchers view (1) as describing a causal relation between two events, so that John finding out causes him to get furious. From this view, (1) should be interpreted probabilistically, in terms of the strength of the evidence one might have for believing that John will get furious given that he finds out. Of course, different people may know different things about John, which may lead them to different evaluations of the probability that he gets furious when he finds out. This could lead to individual differences in reasoning, and as we will see in Section 6.2, such differences have also been observed.

A further theme of this chapter arises from the fact that the experimental work on human reasoning shows that, according to logic, people do make many errors. If being logical is what we mean by being rational then these results may have some serious consequences. For example, in law people can only be held responsible for their actions if they can rationally evaluate the consequences of those actions. But if people are not rational, then how can society hold them responsible for what they do? Moreover, where is the boundary between sanity and insanity to be drawn? A sane person is one who responds rationally to the world and to other people. But if most normal adults are irrational, then who is sane and who is insane? In evaluating psychological theories in this area, it will therefore be important for us to pay close attention to what they have to say about human rationality.

The first issue we address is the sheer ubiquity of human reasoning. Reasoning, like doing the crossword, may seem like an activity we rarely – if ever – engage in. However, most of our common-sense psychological explanations of each other's behaviour assume that we are reasoning all the time.

## 1.2 Reasoning in everyday life

People are so dependent on reasoning processes that they tend to go unnoticed. Nonetheless, it is easy to show that a great deal of human behavior depends on reasoning processes. Suppose you see your neighbour arriving home. She passes her garage and sees that her partner's car is on the drive. When she reaches the door, instead of taking out her key and opening the door as she has done every night for the last twenty years, she rings the doorbell. Why your neighbour broke her habitual pattern of behaviour can be explained in an instant:

She saw that the car was there. She *knows* that, for it to be there, someone must have driven it from town where her partner dropped her off in the morning. Because she *knows* that only her partner has the keys, she *infers* that her partner drove it. She further *infers* that if the car is on the drive then her partner is in the house and hence he can open the door when she rings the bell. Consequently, rather than take out her key, she rings the doorbell.

Two pieces of information are already given. The first comes from prior knowledge: *if the car was there someone drove it*. The second comes directly from perception, that is, *the car was there*. These two pieces of information are combined in an inference to yield new information: someone drove the car. We can depict the inference as follows.

4

	If the car is there then someone drove it. The car is there.	}	(premises)
Therefore,	Someone drove the car.		(conclusion)

The given information can be regarded as the premises and the new information the conclusion of a passage of reasoning. The subsequent steps that lead your neighbour to the final conclusion that she should ring the doorbell can all be characterized in the same way. So it would seem that even the most mundane passage of human behaviour involves complex reasoning processes that require using given information (premises) to infer new information (conclusions).

So, how do we know when we can draw a conclusion from a set of premises? Suppose we had replaced the conclusion in example (4) above with the statement ‘Her partner had a cream tea’. This would be new information too, but the conclusion does not seem to be valid – it does not seem to be related to the premises in the right kind of way. Simply believing the premises in (4) does not compel you to believe this conclusion. Trying to describe the relationship between premises and valid conclusions is the core of characterizing *logical or deductive* reasoning. The idea, as with Aristotle and Boole, is to provide rules that indicate when a conclusion does or does not follow from a set of premises. This is the subject of Section 2.

## Summary of Section 1

---

- Human reasoning and rationality have been understood, contrastively, in terms of the use of logic or probability.
- Human reasoning, involving the drawing of new information (conclusions) from given information (premises), is ubiquitous.

## 2 Deductive reasoning and logic

Logic provides an account of the relationship between the premises and the conclusion of a deductive argument. First, we shall look at the structure of a logical argument.

### 2.1 Logical connectives

Logic starts from the idea that sentences are made up of two very general building blocks. First, there are **descriptive clauses** or sentences that say something true or false about the world, for example, ‘John has a runny nose’ or ‘John has a cold’. Second, there are **structure-building words** that allow us to combine sentences to produce more complex sentences, for instance, ‘John has a runny nose *and* John has a cold’, ‘*if* John has a runny nose, *then* he has a cold’. Along with *and* and *if ... then*, other structure-building words include *or* and *not*. Collectively, these words are called **connectives** because they connect two simpler sentences together.

The most important of these is *if ... then*, which forms, schematically, sentences of the form *if p then q*, which are called **conditionals**. The *p* clause is called the **antecedent** and the *q* clause is called the **consequent**.

### 2.2 When are arguments logically valid?

In a **logically valid** argument, the truth of the premises guarantees the truth of the conclusion. Let us assume that sentences are simply true or false depending on whether what they say really is the case. So the sentence ‘John has a runny nose’ is true if and only if John has a runny nose; otherwise, it is false. How do we determine whether complex sentences made using the connectives are true or false? Well, *not* is particularly simple because all it does is reverse the truth value of a proposition. So, if ‘John has a runny nose’ is true then ‘John does not have a runny nose’ must be false. Conversely, if ‘John has a runny nose’ is false then ‘John does not have a runny nose’ must be true.

What about the other connectives? You should note that the other three connectives all connect two sentences, for example, ‘John has a runny nose *and* John has a cold’. Each sentence making up this complex sentence could be either true or false. Thus, there are four possibilities. If *p* = ‘John has a runny nose’ and *q* = ‘John has a cold’, then

- either *p* is true and *q* is true
- or *p* is true and *q* is false
- or *p* is false and *q* is true
- or *p* is false and *q* is false.

In only one of these possibilities would we intuitively say that the complex sentence ‘John has a runny nose *and* John has a cold’ is true, that is, when *p* is true and *q* is true; otherwise, this complex sentence is false. So we can regard the connective *and* as mapping pairs of truth values on to a truth value. When both *p* and *q* are true, then *p and q* is true; for all other pairs, it is false. We can show this mapping in what is called a ‘truth table’ (see Table 12.1).

**Table 12.1** Truth tables for the logical connectives *not*, *and*, and *if ... then*. The second and third columns show the four possible combinations of truth values for sentences  $p$  and  $q$ . Each is assumed to be either true or false. The four columns to the right show for each connective the truth value of the complex sentence formed from the connective and the sentences  $p$  and  $q$

Possibility	Sentences		Connectives			
	$p$	$q$	not	and	if ... then	
			not $p$	not $q$	$p$ and $q$	
A	true	true	false	false	true	true
B	true	false	false	true	false	false
C	false	true	true	false	false	true
D	false	false	true	true	false	true

Now try Activity 12.1

### ACTIVITY 12.1

Use Table 12.1 to see whether you agree with the truth tables for *not* and for *and*. Choose any two simple sentences (i.e. ones not already containing a connective such as *or*, *if ... then*, *or and*). Let  $p$  stand for one of the sentences and  $q$  for the other. Then consider the four possibilities A to D in turn (e.g. in A,  $p$  is true and  $q$  is true). That is, imagine your two sentences are either true or false, according to each possibility. Then for each possibility try to work out what you would say about the truth of the two sentences (1) *not p* and (2) *p and q*.

Truth tables such as those in Table 12.1 illustrate what we have said about the relation between the premises and conclusion of a deductive argument. We said that if you believe the premises of a deductive argument, then somehow you are compelled to believe the conclusion. This means, for example, that if it is (always) true that ‘if John has a runny nose then he has a cold’ and it is true that ‘John has a runny nose’, then it has to be true that ‘John has a cold’. Table 12.1 shows this. Again, let  $p$  = ‘John has a runny nose’ and  $q$  = ‘John has a cold’. If we look at Table 12.1, then we see that whenever *if p then q* and *p* are true (i.e. possibility A), *q* is also true. So if both premises *if p then q* (the conditional) and *p* (the antecedent) are true, then the conclusion *q* (the consequent) must be true – there are no other possibilities! We describe this by saying that the inference, or the drawing of this conclusion, is logically valid. This particular form of inference is also referred to as **modus ponens (MP)** and can be depicted as follows:

5	Modus Ponens (MP)	
	Inference in schematic form	Example
	If $p$ then $q$	If John has a runny nose, then he has a cold
	$p$	John has a runny nose
	∴ $q$	Therefore, John has a cold

Another logically valid inference is **modus tollens (MT)** for short, illustrated below.

6	Modus Tollens (MT)	
	Inference in schematic form	Example
	If $p$ then $q$	If John has a runny nose, then he has a cold
	Not $q$	John does not have a cold
	∴ Not $p$	Therefore, John does not have a runny nose

We can similarly show that this inference is logically valid using Table 12.1. Of course, *not q* is true when *q* is false. Now, *if p then q* and *not q* are both true only in the last line of the truth table (possibility D). However, in possibility D, *p* is false, and so *not p* is true. So, if the two premises are true, then *not p* must be true also, as again there are no other possibilities. So if the conditional is true and John does not have a cold, then John cannot have a runny nose.

### 2.3 Logically invalid inferences

Two logically invalid inference patterns have been investigated in the psychology of reasoning: **affirming the consequent (AC)** and **denying the antecedent (DA)**:

7	Affirming the consequent (AC)	
	Inference in schematic form	Example
	If $p$ then $q$	If John has a runny nose, then he has a cold
	$q$	John has a cold
	∴ $p$	Therefore, John has a runny nose

8	Denying the antecedent (DA)	
	Inference in schematic form	Example
	If $p$ then $q$	If John has a runny nose, then he has a cold
	Not $p$	John does not have a runny nose
	∴ Not $q$	Therefore, John does not have a cold

Remember, both of these forms of inference are *not* logically valid, even though you may feel that they appear to make some sense. To see why they are not logically valid, try Activity 12.2.

## ACTIVITY 12.2

---

Using Table 12.1, try to work out why AC and DA are not logically valid. Remember how we saw that MP and MT are valid. First, we worked out in which of the four possibilities A to D both premises were true at the same time. Second, we considered whether the conclusion was always true in those possibilities. For logically invalid inferences, there should be a possibility in which both premises are true but the conclusion is false. Which possibilities show AC and DA to be logically invalid? Answers are given at the end of the chapter.

---

## 2.4 Form and meaning in logic

We can now illustrate a critical distinction in logic between *form* and *meaning*. Examples (5) and (6) above show the *form* of the logically valid inferences MP and MT. We know they are logically valid because of Table 12.1. The table encodes the *meanings* of the connectives. It tells us that if the premises are true then the conclusion must be true (which, as you saw above, is the definition of logical validity). However, to draw the MP or MT inference we do not need to make reference to the truth table, that is, to the meaning of the conditional.

To draw the inference in (4), for example, all you need do is match this example to the *formal rule* in (5). (4) is simply a particular instantiation of (5). Using (5) we can automatically make the logically valid inference, simply because of the *form* of the argument, without worrying about what it means. This distinction between *form* and *meaning* is central to two of the theories of reasoning we look at later on. One theory – **mental logic** – argues that we have formal inference rules, like (5), in our heads, so that drawing inferences relies only on *form*. The other, **mental models**, argues that we do something much more like considering Table 12.1, so that drawing inferences relies on the *meaning* of the connectives.

## Summary of Section 2

---

- Logic provides rules, based on the truth or falsity of propositions, to determine whether an inference is valid.
- A (logically) valid inference is one in which the conclusion is always true whenever the premises are true.
- There are two ways of establishing the validity of an inference: semantic, using truth tables, and formal, using rules of inference (e.g. (5) and (6)).

### 3 Psychological theories of reasoning

There are a wide variety of theoretical approaches to the psychology of human reasoning. In Section 3, the main theoretical approaches are discussed: mental logic, mental models and the probabilistic approach. These are all *general* theories of reasoning in that they are intended to apply to most reasoning tasks.

#### 3.1 Mental logic

The **mental logic** group of theories (there are several different versions of the basic account [Braine and O'Brien, 1998; Rips, 1994]) are also known as **formal rule** theories. As the name suggests, these accounts are close in spirit to Piaget's view that adult human thought is the operation of formal logic (Inhelder and Piaget, 1958). The idea behind these theories is that people possess a system of formal mental logic that contains inference rules such as (5) above. However, people's failure to reason logically all of the time can then be explained by assuming that they do not possess all the formal logical rules that are licensed by the truth tables in Table 12.1. Without a particular rule, some inferences will be more difficult to make than others.

#### 3.2 Mental models

Mental models theory (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991) shares the intuition with mental logic that people are in principle capable of logical reasoning. However, rather than applying formal rules, mental models theory argues that people reason over pictorial representations of what sentences *mean*. These representations concern the different *possibilities* that a logical expression may allow (just as each row in Table 12.1 concerns a different possibility).

One way of thinking about the connectives is that they exclude different possibilities. For example, if *if p then q* is true then there are only three possibilities (A, C and D in Table 12.1) – it would not be possible for *p* to be true and *q* false (B). So, to interpret the sentence *if p then q* people may need to hold in mind three possibilities – that is, they may need a mental model that represents each possibility. However, given the limited capacity of working memory (as you saw in Chapter 9), people may not be able to represent all of these possibilities at once. Rather, there may be a preferred initial representation or interpretation. This idea is the core of the mental models theory. In this account each possibility is referred to as a 'mental model'. How people manipulate these mental models explains their reasoning performance.

#### 3.3 The probabilistic approach

According to the probabilistic approach, logic simply does not provide the right framework for understanding people's everyday inferences (Oaksford and Chater, 1994, 1998). For example, perversely, according to logic, a good reason to believe that *if John has a runny nose then he has a cold* is that you do *not* believe that *John has a runny nose*. This is because, according to logic, a conditional is true whenever its antecedent is false (see possibilities C and D in Table 12.1). But *not* believing that *John has a runny nose* is not sufficient grounds for believing the conditional. What appears to be required is the belief that John's having a runny nose makes it very likely that he has a cold. This involves assessing the *conditional probability* that

John has a cold given that he has a runny nose. So if you have noticed John having a runny nose on say 100 occasions, 95 of which involved him having a cold, then the relevant conditional probability is 0.95. We can write this as follows:

$$P(\text{John has a cold} \mid \text{John has a runny nose}) = 0.95 \\ (\mid \text{should be read as 'given'})$$

This implies that belief in the conditional is a matter of degree, rather than the (completely) true or (completely) false implied by logic. This was recently confirmed by Evans *et al.* (2003). If all you know is that John has a runny nose, then your degree of belief in this rule indicates that there is a 95 per cent chance that he has a cold. According to the probabilistic approach, most inferences are *uncertain* because the conditional rules on which they are based describe the real world in which logical certainty is a rare commodity. This account of reasoning performance therefore suggests replacing logic with probability theory as the framework for understanding inferences people should make.

The contrast between the first two general theories of reasoning and the last is important because the concept of what it is to be rational fundamentally changes. In the first two theories, rationality is still defined as logical reasoning, and errors (or apparent departures from rationality) are explained in terms of performance limitations such as the limited nature of short-term memory. In contrast, according to the probabilistic approach, probability theory replaces logic as the criterion of what is rational, and rationality is defined in terms of probabilistic reasoning.

However, assessing theories in any area of science depends first and foremost on the ability to explain data. In Sections 4 and 5 we look at some of the experimental results from the two principal reasoning tasks that have been used to assess human reasoning: conditional inference and Wason's selection task.

## Summary of Section 3

---

- Mental logic theories assume people use formal rules of inference but errors may arise because they do not possess all of the possible rules.
- Mental models theory assumes people represent the true possibilities licensed by connectives. Failing to represent all of these possibilities may then lead to errors.
- The probabilistic approach assumes people are not drawing logical, deductive inferences but endorse inferences based on their assessment and evaluation of appropriate conditional probabilities.

## 4 Conditional inference

In a **conditional inference** task (Evans, 1977; Taplin, 1971) participants are presented with a conditional sentence (the conditional premise) and various facts relating either to the antecedent or consequent of the sentence (the categorical premise). In our example, the conditional premise would be *if John has a runny nose, he has a cold*. Different categorical premises would then relate to different schemes of inference: for example, *John has a runny nose* relates to the MP inference, *John does not have a cold* relates to MT, *John has a cold* to AC and *John does not have a runny nose* to DA. Participants are asked to indicate what conclusion follows from these two premises, that is, the categorical and the conditional premises. Logic dictates that inferences should be made for the logically valid schemes of inference MP and MT, and withheld for the logically invalid inferences AC and DA. We now review the main findings on conditional inference. For each finding we then discuss how the main theories explain the results.

### 4.1 The abstract conditional inference task

Typically, these reasoning tasks are conducted using abstract alphanumeric stimuli – letters and numbers – in order to try and rule out any effects of prior knowledge and so to investigate reasoning ‘in the raw’, that is, to investigate the basic operating characteristics of the cognitive system.

In the abstract conditional inference task participants are told to assume that, for example, the premises, *if there is an A then there is a 2* and *there is an A*, are true and are asked whether they can conclude that *there is a 2* (as MP indicates). Schroyens and Schaeken (2003) summarized 65 of these experiments, and their results are shown in Table 12.2.

**Table 12.2** Proportion of participants endorsing inferences in the abstract conditional inference task

Inference type	Proportion of participants endorsing inference (%)
MP	97
MT	72
AC	63
DA	55

Source: adapted from Schroyens and Schaeken, 2003

Comparing the mean values in Table 12.2 pair-wise, all six comparisons reveal highly significant differences. So, MP is endorsed more than MT, which is endorsed more than AC, which is endorsed more than DA. This pattern is not consistent with logic, according to which participants should endorse MP and MT fully and equally, and not endorse other inferences.

### 4.1.1 Mental logic

Mental logicians (e.g. Rips, 1994) explain the difference in the extent to which people endorse MP and MT by proposing that people possess the MP inference rule (5) but do not possess the MT inference rule (6). In order to draw the MT inference, much more complex reasoning would be needed, as outlined below in (9).

Reconsider the example of MT given in (6).

- 9      *If John has a runny nose, then he has a cold*  
*John does not have a cold*  
*Therefore, John does not have a runny nose*

To reach the conclusion, Rips (1994) argues that people assume the contrary is true, that is, they assume that *John has a runny nose*, and then find that a contradiction results. If *John has a runny nose* is true then, using MP, it can be combined with the conditional premise to yield the conclusion *John has a cold*. But this conclusion contradicts the actual categorical premise *John does not have a cold*. Since the assumption results in a contradiction, the original assumption must be false, that is, *John has a runny nose* must be false and so *John does not have a runny nose* must be true! This way of drawing the MT inference is called **reductio ad absurdum** or RAA. The complexity of this inference is then thought to explain why the MT inference is drawn less often than MP.

Indeed, this strategy may also explain the results for DA and AC (Rips, 1994). Conditionals in natural language can be ambiguous and may sometimes be interpreted as **bi-conditionals**. Table 12.3 gives a truth table for a bi-conditional, sometimes expressed as *if and only if ... then ...*

**Table 12.3** A truth table for the bi-conditional *if and only if ... then ...*. Note how the bi-conditional *if and only if p then q* is true only when both *if p then q* and *if q then p* are true

<b>Possibility</b>	<b>Sentences</b>		<b>Connectives</b>		
	<b>p</b>	<b>q</b>	<b>Complex sentences</b>		
			<b>If p then q</b>	<b>If q then p</b>	<b>If and only if p then q</b>
A	true	true	true	true	true
B	true	false	false	true	false
C	false	true	true	false	false
D	false	false	true	true	true

As Table 12.3 shows, the bi-conditional *if and only if John has a runny nose, then he has a cold* is true whenever both the standard conditional *if John has a runny nose, then he has a cold* and the converse conditional *if John has a cold, then he has a*

*runny nose* are true. This means that if people interpret *if John has a runny nose, then he has a cold* as a bi-conditional, then they should draw the logically valid MP and MT inferences on both the standard conditional (*if p then q*) and also on the converse conditional (*if q then p*). However, note that MP and MT on the converse conditional are equivalent to AC and DA on the standard conditional: that is MP on *if John has a cold, then he has a runny nose* is equivalent to AC on *if John has a runny nose, then he has a cold*; similarly, for MT and DA. Consequently, those participants who interpret the rule as a bi-conditional should endorse both DA and AC. However, since AC is equivalent to the easier MP on the converse conditional, and DA to the more difficult MT, they should endorse AC more than DA. This corresponds well to the pattern of endorsements observed in Schroyens and Schaeken's (2003) summary of the data. So, according to mental logic theory, people's performance on this task is rational because it is actually logical: it is just that (1) some logical inferences are harder than others given the logical rules we possess and (2) some people misinterpret the conditional as a bi-conditional.

#### 4.1.2 Mental models

Table 12.4 (a) shows the **initial mental model** representation for the conditional *if p then q*. It represents the possibility in which *p* is true and *q* is true (like possibility A in Table 12.3). The three dots (or ellipsis) indicate that there may be other relevant mental models or possibilities that are not currently being considered. These other possibilities are available, perhaps temporarily held in some short-term memory store, but are not explicitly represented. The square brackets indicate that *p* cannot be paired with any other term. In particular, this captures the fact that *p* cannot be paired with *not q*, because this is the possibility that the conditional excludes (possibility B in Table 12.1). Table 12.4 (c) shows the initial mental model for the bi-conditional *if and only if p then q*. Both *p* and *q* are now in square brackets because neither can be paired with anything else: this rules out the *p, not q* (possibility B) and the *not p, q* (C) possibilities.

**Table 12.4** Initial mental models for the conditional (a) and bi-conditional (c), and fleshed-out versions of these (b) and (d) respectively. The three dots (ellipsis) indicate that there may be other relevant mental models not explicitly represented. The square brackets indicate that an item cannot be paired with any other term. The inferences that can be drawn from each model are abbreviated at the bottom of the table

Conditional <i>if p then q</i>		Bi-conditional <i>if and only if p then q</i>			
Initial model	Fleshed-out model	Initial model	Fleshed-out model		
[p]      q ... (a) MP	[p]      q not p      q not p      not q (b) MP MT	[p]      [q] ... (c) MP AC	[p]      [q] not p      not q (d) MP DA AC MT		

Table 12.4 (b) shows the **fleshed-out mental model** for the conditional, where the other possibilities not excluded by the initial interpretation in Table 12.4 (a) are now explicitly represented. Table 12.4 (d) similarly shows a fleshed-out mental model for the bi-conditional. Now try Activity 12.3.

### ACTIVITY 12.3

---

For each of the inferences MP, MT, AC and DA, using the mental models shown in Table 12.4, try to work out why each inference can be made (or not as the case may be) in each model. You should find that the process is very like checking the truth table in Activity 12.1.

#### COMMENT

Suppose participants adopt the initial conditional interpretation in Table 12.4 (a). While the categorical premise of both MP ( $p$ ) and AC ( $q$ ) would match the model,  $q$  could be paired with something other than  $p$ . So, given the categorical premise  $q$ , no conclusion can be drawn. However,  $p$  can only be paired with  $q$  (that is what the square brackets mean) so, given the categorical premise  $p$ , participants can conclude  $q$ . That is, they will only be able to draw the MP inference.

Participants that adopt the fleshed-out conditional interpretation in Table 12.4 (b) can make both the MT and MP inferences. Although all categorical premises now match the model, only  $p$  and *not q* are constrained in their pairings. From  $p$  we can conclude  $q$  (MP) since  $p$  is only paired with  $q$ , and from *not q* we can conclude *not p* (MT) since *not q* is only paired with *not p*.

Participants who adopt the initial bi-conditional interpretation in Table 12.4 (c) can make the MP inference and the AC inference. MP goes through for the same reason as for the initial conditional interpretation. AC goes through because now  $q$  can only be paired with  $p$ .

Participants that adopt the fleshed-out bi-conditional interpretation in Table 12.4 (d) can draw all inferences. This is because the categorical premise of all inferences find a match in the model and each is uniquely paired with only a single item.

---

The fact that mental models only represent the true possibilities, together with the distinction between initial and fleshed-out mental models, is the primary means by which mental models theory explains the data on conditional inference. The explanation depends on assuming different subsets of participants adopt these four different mental models. Table 12.4 shows that MP can be endorsed in all representations, DA can only be endorsed in one, and AC and MT can both be endorsed in two. So if equal numbers of participants adopted each representation, then MP would be endorsed more than both AC and MT, which would be endorsed in equal proportion, and all would be endorsed more than DA. That participants actually endorse MT more than AC can be explained by assuming that more participants flesh out the conditional interpretation in Table 12.4 (b) than adopt the initial bi-conditional interpretation in Table 12.4 (c).

### 4.1.3 Probabilistic approach

According to the probabilistic approach people draw inferences according to how probable they think the conclusion is given the premises (Oaksford and Chater, 2003a; Oaksford *et al.*, 2000). MP is straightforward. Given the conditional premise, *if John has a runny nose, then he has a cold*, this inference will be drawn in proportion to the conditional probability that *John has a cold* given that *he has a runny nose*, a probability that we can write as  $P(\text{cold} | \text{runny nose})$ . So for example, given the categorical premise, *John has a runny nose* (and no other information) the best bet as to the probability that *John has a cold* is the proportion of times that John has a cold when he has a runny nose. In our example above, we supposed this to be 0.95. Given that the consequent is highly likely given the antecedent, we can assume it is equally highly likely that people will endorse this inference. Calculating probabilities for the remaining inferences requires the assumption that people possess information about the probability of John having a runny nose ( $P(\text{runny nose})$ ) and the probability that he has a cold ( $P(\text{cold})$ ). Consider the AC inference in which you are told that *John has a cold* and are asked whether you would endorse the conclusion that *John has a runny nose*. What you are interested in is the probability that *John has a runny nose* given *John has a cold* ( $P(\text{runny nose}|\text{cold})$ ), i.e., the converse of MP. Probability theory allows us to calculate this probability using a form of Bayes' Theorem (note that Chapter 11 used a somewhat different form):

$$P(\text{runny nose}|\text{cold}) = \frac{P(\text{cold}|\text{runny nose})P(\text{runny nose})}{P(\text{cold}|\text{runny nose})P(\text{runny nose}) + P(\text{cold}|\text{not runny nose})P(\text{not runny nose})}$$

We know that  $P(\text{cold}|\text{runny nose}) = 0.95$ . Suppose that John rarely has a runny nose and so over the year he has only a 5 per cent chance of having one, so  $P(\text{runny nose}) = .05$  and therefore  $P(\text{not runny nose}) = 0.95$ . What about  $P(\text{cold}|\text{not runny nose})$ , that is, the probability of *John having a cold* given *he does not have a runny nose*? This is likely to be quite low but not 0. Let us set this to .03.  $P(\text{runny nose}|\text{cold})$  can then be calculated using the equation above and comes to 0.61. This then would be the probability of drawing the AC inference. Probabilities can be derived in a similar way for DA and MT.

To assess how well this account can explain the standard abstract results, the equations predicting the probabilities with which each inference should be drawn can be fitted to the data. This means that the values of  $P(\text{cold}|\text{runny nose})$ ,  $P(\text{runny nose})$ , and  $P(\text{cold})$ , which we chose freely in the above example, are chosen to provide the best possible predictions for the frequencies with which each inference is endorsed. The probabilistic account provides a close fit to the abstract data, predicting the following frequencies for each inference (actual frequencies in brackets): MP = 0.88(0.97), DA = 0.51(0.55), AC = 0.68(0.63), MT = 0.77(0.72) (Oaksford and Chater, 2003a). Note that fitting an account to the data in this way is not unique to the probabilistic approach. The numbers of participants adopting the conditional or bi-conditional interpretation and the numbers fleshing out or drawing RAA inferences are all free to vary and must be fixed from the data, just as probabilities are in the probabilistic approach.

## 4.2 Everyday reasoning and the suppression effect

A general property of everyday inferences is that they can be defeated (Oaksford and Chater, 1998). For example, if you infer that *John has a cold* because *he has a runny nose*, but then discover that *he has hay fever*, your inference is defeated. Because such inferences can be defeated in this way they are called **defeasible inferences**. Take another example: we generally believe that *birds fly*, which can be expressed as the conditional *if something is a bird, then it flies*. So we might conclude that *Tweety can fly* on learning that *Tweety is a bird*. However, if we then discover the new information that *Tweety is an ostrich*, then this inference is defeated. It thus seems that many of the inferences we draw in everyday life may be non-deductive, that is, the truth of the conclusion is not guaranteed by the truth of the premises.

There have been many experiments investigating these aspects of everyday reasoning. They show that the inferences, MP and MT, and the fallacies, DA and AC, can be **suppressed** by providing information about possible defeaters. For example, if you are told that *if the key is turned then the car starts* and that *the key is turned*, you are likely to endorse the MP inference to the conclusion that *the car starts*. However, if you are also told that *the petrol tank is empty*, you are less likely to endorse this conclusion because the car will *not start* if the petrol tank is empty. An empty petrol tank provides an *exception* to the rule. This exception would also mean that you are less likely to endorse MT. If you knew that *the car didn't start* you may not infer that *the key was not turned* because the empty petrol tank may be the cause of the car not starting. These exceptions have been called 'additional antecedents' (Byrne, 1989).

Other information can suppress DA and AC. For example, if you are told that *if the key is turned then the car starts* and that *the key is not turned*, you might be tempted to endorse the DA inference to the conclusion that *the car does not start*. However, if you are also told that *the car was hot-wired*, you may be less likely to endorse this conclusion (because the car may start even though the key was not turned because it has been hot-wired). This condition would also mean that you are less likely to endorse AC. If you knew that *the car started* you may not infer that *the key was turned* because the car starting may have been caused by being hot-wired. These conditions have been called 'alternative antecedents' (Byrne, 1989).

Byrne (1989) demonstrated all these effects by providing participants with explicit rules containing this additional information, as in (10) and (11) below.

10

### Additional antecedents (MP)

*If the key is turned the car starts*

*If there is fuel in the tank the car starts*

*The key is turned*

*The car starts?*

II Alternative antecedents (AC)

If the key is turned the car starts

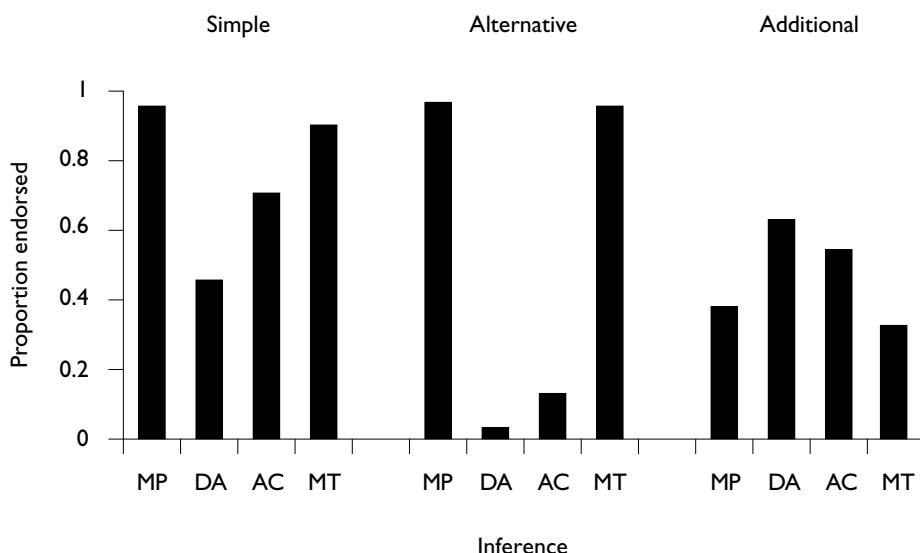
If it is hot-wired the car starts

The car starts

The key was turned?

The results of Byrne's (1989) Experiment 1 are shown in Figure 12.1 below. The simple condition did not include any additional or alternative antecedents and reflects the standard pattern of results (see Table 12.2). Figure 12.1 shows clearly that providing participants with information concerning alternative antecedents suppresses DA and AC but not MP or MT, whereas information concerning additional antecedents suppresses MP and MT but not DA or AC.

Similar effects have been demonstrated *without* presenting additional or alternative antecedents explicitly (Cummins *et al.*, 1991; Cummins, 1995). Thus, people seem to automatically retrieve this information from memory to influence their reasoning performance. A range of causal conditionals, like those we have looked at, were pre-tested for the number of additional or alternative antecedents participants could bring to mind. It was shown that the effects of additional or alternative antecedents were *graded*, that is, the more additional or alternative antecedents a rule allowed the greater the suppression effects observed. As we will see, this result does not sit well with theories that regard these effects as *all or nothing*. In experiments where participants *rated* how likely an inference was to go through on a scale of 1–7 (Cummins *et al.*, 1991; Cummins, 1995) *all or nothing*



**Figure 12.1** Suppression effects

Source: Byrne, 1989, Table I

behaviour suggests the data should be made up of different proportions of two types of individual: those who give only ratings of 1, meaning they ‘do not endorse’ the conclusion, and those who give only ratings of 7, meaning they do ‘endorse’ the conclusion. However, in experiments like these, around 50 per cent of each participant’s responses are *intermediate* values (i.e. not ratings of 1 or 7) (Oaksford *et al.*, 2000, Experiment 3), which is not consistent with all or nothing behaviour.

#### 4.2.1 Mental logic

In Byrne’s (1989) experiments (see (10) and (11) above) participants were provided with two conditional rules with the same consequents. So for additional antecedents, (10), the key needs to be turned and there needs to be fuel in the tank for the car to start. However, for alternative antecedents (11) either turning the key *or* hot-wiring will start the car. This means that people may represent these premises as single rules with complex antecedents.

10'      *If the key is turned and there is fuel in the tank, then the car starts*

11'      *If the key is turned or it is hot-wired, then the car starts*

Let us look at the MP and MT inferences for (10').

- MP inference: the categorical premise *the key is turned* does not satisfy the antecedent, which is only true if *the key is turned and there is fuel in the tank* (see Table 12.1). Consequently, logically this single categorical premise does not license the conclusion that *the car starts* and so the MP inference cannot be drawn.
- MT inference: from the categorical premise *the car does not start*, by MT, one can conclude that the complex sentence *the key is turned and there is fuel in the tank* is false. Logically, one cannot infer that *the key was not turned* because there could be no fuel in the tank, and so the MT inference should not be drawn. Similar reasoning applies to (11') for the DA and the AC inferences.

#### ACTIVITY 12.4

---

For (11') see if you can work out why DA and AC should be suppressed. A clue for the DA inference is that *not(the key is turned or the car is hot-wired)* is equivalent to *the key is not turned and the car is not hot-wired*. Table 12.5 gives the truth table for *or*.

**Table 12.5** A truth table for or

<b>Possibility</b>	<b>Sentences</b>		<b>Connective or</b>	<b>Complex sentence</b>
	<b>p</b>	<b>q</b>	<b>p or q</b>	
A	true	true		true
B	true	false		true
C	false	true		true
D	false	false		false

Although mental logic can offer explanations for suppression effects, it nevertheless suggests that reasoning behavior should be all or nothing – that conclusions should either be endorsed or not. This is hard to reconcile with Cummins' (1995) data, which revealed graded effects. One possible response is to suggest that mental logic provides a good explanation of deductive inference but that graded effects tap non-deductive inference (Rips, 2001, 2002b) (see Section 6.1.1).

#### 4.2.2 Mental models

The mental models explanation of the suppression effects depends on the availability of counter-examples. Just as for the mental logic approach, mental models theory (Byrne *et al.*, 1999) suggests that people represent information about additional antecedents using *and*, and they represent information about alternative antecedents using *or*. This yields initial mental models representations for (10') and (11') as shown in the top part of Table 12.6.

**Table 12.6** Mental models representations for (10') and (11')

<b>Additional antecedents</b>			<b>Alternative antecedents</b>		
<i>If the key is turned and there is fuel in the tank then the car starts</i>			<i>If the key is turned or it is hot-wired then the car starts</i>		
turn	fuel	starts	turn		starts
				hot-wired	starts
...			...		
turn	not (fuel)	not (starts)	not (turn)	hot-wired	starts
	(10')			(11')	

For (10') a fully fleshed-out version of the mental model for *and* will include the case where the key is turned but the car does not start because there is no fuel. For (11') a fully fleshed-out version of the mental model for *or* will include the case where the key is not turned but the car starts because it was hot-wired. These models are shown after the ellipsis. This particular example shows why these counter-examples need to be available.

We have deliberately picked an example that appeals to your prior knowledge of cars and the factors that determine whether they start or not. This is called the **principle of pragmatic modulation** (Johnson-Laird and Byrne, 2002). That is, general knowledge in long-term memory can modulate the interpretation of

conditionals, in this case making certain counter-examples much easier to access and represent.

Mental models also suggests that suppression effects should be all or nothing. However, some researchers working within the mental models framework (Quinn and Markovits, 2002; Schroyens and Schaeken, 2003) have suggested that mental models should be supplemented with a validating search procedure. A conclusion is suggested by the mental model, and then long-term memory is searched to see if there is a counter-example. These might influence reasoning either in an all or nothing way (Quinn and Markovits, 2002) or in a graded, probabilistic way (Schroyens and Schaeken, 2003). Opting for this explanation of suppression effects means that people can no longer be thought of as performing strictly logical inferences. If people take a conditional to be true then, logically speaking, there can be no need to search for counter-examples – people should only search long-term memory for counter-examples if they are not strictly taking the conditional to be true.

#### 4.2.3 Probabilistic approach

Suppression effects in conditional inference are explained in terms of the effects of additional and alternative antecedents on the appropriate conditional probabilities (Oaksford and Chater, 2003c). Suppose you were asked to estimate the probability of a car starting given you have turned the key. You might base your estimate on the proportion of times cars have started when you have turned the key. Suppose you are now provided with an *additional antecedent*, or you retrieve one from memory – perhaps the possibility that the petrol tank is empty. Now estimate again the probability of the car starting given you have turned the key. As an empty petrol tank will prevent the car from starting, presumably this probability will now be smaller than in your first estimate, when all you were told was that the key was turned. That is, your estimate of the probability that the car starts is suppressed. In the suppression experiments people are not provided with the information in this way. Rather, they are given reminders about general preventative factors, like empty fuel tanks. According to the probabilistic approach, this has the effect of reducing people's estimates of the conditional probability of the car starting given you turn the key,  $P(\text{car starts}|\text{key turned})$ . Thus, information about additional antecedents suppresses MP and MT inferences.

Explaining suppression effects for alternative antecedents follows a similar pattern. Alternative antecedents emphasize that, for example, it is possible to start cars without turning the key, for instance, by hot-wiring. They therefore suggest that the probability of the car starting given you don't turn the key,  $P(\text{car starts}|\text{key not turned})$ , is higher than you first thought. This has to reduce the probability of the car not starting given that you do not turn the key,  $P(\text{car does not start}|\text{key not turned})$ . This is simply because these probabilities must sum to 1, so that,  $P(\text{car starts}|\text{key not turned}) + P(\text{car does not start}|\text{key not turned}) = 1$ . So, if one goes up, the other must come down. The probability of the car not starting given that you do not turn the key is the probability that you must assess to determine whether to draw the DA inference. This is why alternative antecedents suppress the DA and AC inferences according to the probabilistic approach.

## Summary of Section 4

---

- Both the standard abstract task and suppression experiments yield results that appear inconsistent with logic.
- Mental logic explains these by assuming that people use a more complex inference to perform MT, and that some adopt a bi-conditional interpretation. Additional and alternative antecedents are represented in terms of complex antecedents (using the connectives *and* and *or*).
- Mental model theory assumes some people adopt the bi-conditional interpretation and that some do not flesh out initial semantic representations. Suppression effects are explained by the principle of pragmatic modulation.
- The probabilistic approach suggests that people reason by judging the probabilities and conditional probabilities of events. Suppression effects arise because additional and alternative antecedents modulate these probabilities.
- Graded effects, which can be explained by the probabilistic approach, have led some mental logicians to argue for a distinction between deductive and non-deductive inference and some mental models theorists to introduce a probabilistic component.

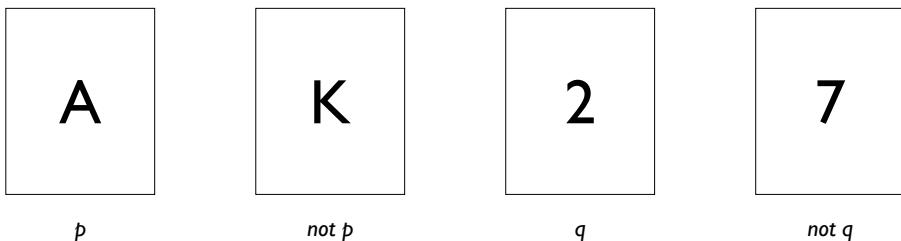
## 5 Wason's selection task

Wason's selection task is probably the most used task in the psychology of reasoning (Wason, 1968). We look first at the original *abstract* form of the task and then at how each theory explains the data.

### 5.1 The abstract selection task

In this version of the task people assess whether evidence is relevant to the truth or falsity of a conditional rule (Wason, 1968). In the abstract version, the rule concerns cards that have a number on one side and a letter on the other (see Figure 12.2). A typical rule is *if there is an A on one side* (call this *p*), *then there is a 2 on the other side* (*q*).

Four cards are placed before the participant, so that just one side is visible, showing an *A* (*p* card), a *K* (*not p* card), a *2* (*q* card) and a *7* (*not q* card) (Figure 12.2). Participants are told that each card has a letter on one side and a number on the other. They are then asked to pick those cards they must turn over to test whether the rule *if there is an A on one side then there is a 2 on the other side* is true or false. It was shown in Table 12.1 that sentences like *if p then q* are only false when the antecedent, *p*, is true (there is an *A* on one side) and the consequent, *q*, is false (there is not a 2 on the other side). Consequently, according to logic, only a card with an *A* on one side but without a 2 on the other side makes this rule false. There are only two cards that could possibly be of this type: the *A* card could have a number other than 2 on the other side, and the *7* card could have an *A* on the other side. So, logically, people should select the *A* and the *7* cards to turn over, as these are the only cards that could falsify the rule, but not the *K* or the *2* cards.



**Figure 12.2** The four cards in the abstract version of Wason's selection task. For convenience, the cards have been annotated according to their match to *if p then q* (these annotations are not seen by participants)

However, these selections are rarely observed in the experimental results, as you will see in Table 12.7. That is, as few as 4 per cent of participants make the response predicted by logic. Participants typically select cards that could confirm the rule, that is, the  $p$  and  $q$  cards. However, according to logic, the choice of the  $q$  card is irrational, and is an example of so-called **confirmation bias**. That is, people appear to be trying to find cards that have an  $A$  ( $p$ ) on one side and a  $2$  ( $q$ ) on the other side. Consequently, if we judge whether people are rational by a logical standard then these results seem to indicate that people are irrational.

**Table 12.7** Typical results in the Wason selection task

Cards selected	Proportion of participants selecting cards (%)
$p$ and $q$ cards	46
$p$ card only	33
$p, q$ and $\text{not } q$ cards	7
$p$ and $\text{not } q$ cards	4
other card combinations	10

Source: Johnson-Laird and Wason, 1970, Table I

The degree of people's apparent irrationality was illustrated further by work on the **matching effect**. Evans and Lynch (1973) used rules that also contained negations, for example, *if there is an A on one side then there is not a 2 on the other side*. Participants continued to select the  $A$  and the  $2$  card. For this rule these are now the logical responses because a card that does *not* have *not-2* on it must have a  $2$  on it, so  $A$  and  $2$  now correspond to the  $p$  and  $\text{not-}q$  cards. However, if they were showing confirmation bias they should now select the  $A$  and the  $7$  card. Selecting the  $A$  and the  $2$  card for both the standard and the negated rules seems only consistent with **matching bias**. That is, participants are not engaging rationally in the task at all but are simply choosing cards that match the letters and numbers mentioned in the rule.

### 5.1.1 Mental logic

The mental logic approach to the selection task is identical to that taken for the conditional inference task. The account then relies on the following identities:  $\text{MP} = p$  card,  $\text{DA} = \text{not } p$  card,  $\text{AC} = q$  card and  $\text{MT} = \text{not } q$  card. That is, the card sides that participants see are taken to be the categorical premises in conditional inferences. So, for example, given the rule *if A then 2*, deciding to turn

the *A* card (*p*) is equivalent to drawing an MP inference to predict that there is a 2 on the other side. Suppose participants misinterpret the rule as a bi-conditional, so they also believe that *if 2 then A*. Given this interpretation, deciding to turn the *not A* card is like drawing the logically valid MT inference to predict that there is a *not 2* on the other side (which is equivalent to the logically invalid DA on the original rule as stated).

You should note that this account predicts that the proportion of people selecting the cards in the selection task should mirror the proportion of participants endorsing the corresponding conditional inferences. But this is not observed. In the selection task, the *q* card is endorsed more than the *not q* card, yet in the conditional inference task the MT inference is endorsed more than AC.

One reason for this discrepancy between the results on the two tasks may be the way in which the categorical premises are presented in the selection task. Take MT. In the conditional inference task, people see *if A then 2, not 2*, therefore, *not A* and are asked to judge the appropriateness of the conclusion. However, in the selection task, they are told *if A then 2*, are shown a card with 7 on one side, and are given no explicit conclusion. The point is that they have to infer that 7 is an instance of the category of numbers that are *not 2*. Presenting a negated premise in this way has been labelled an **implicit negation** (Evans *et al.*, 1996) and this is always the method used in the selection task. When implicit negations are used in the conditional inference task the typical pattern for the *if A then 2* rule becomes very close to the corresponding pattern of card selections in the selection task, given the identities above (as you will see in Table 12.8).

**Table 12.8** Proportion of participants endorsing inferences in the abstract conditional inference task with implicit negations

Inference type	Proportion of participants endorsing inference (%)
MP	95
MT	58
AC	79
DA	38

Source: Evans and Handley, 1999, Table 5

This account may also explain the matching effect. If, for example, *not 2* is shown on the cards, rather than 7, then the matching effect goes away (Evans *et al.*, 1996). So the matching effect may be a result of having to process implicit negations, rather than an inherent illogicality. In sum, it would appear that people may indeed be drawing conditional inferences in the selection task using the upturned face as the categorical premise.

### 5.1.2 Mental models

Mental models theory also explains the selection task results in a similar way to the conditional inference task. However, now people consider each possibility for whether there could be something on the other side of the card that bears on the truth or falsity of the rule. The frequencies of card selections depend on the proportions

of participants adopting the different interpretations. We examine the card that should be turned for each interpretation by looking at Table 12.9 (which is a copy of Table 12.4).

**Table 12.9** Initial mental models for the conditional (a) and bi-conditional (c), and fleshed-out versions of these (b) and (d) respectively. The three dots (or ellipsis) indicate that there may be other relevant mental models not explicitly represented. The square brackets indicate that an item cannot be paired with any other term. The inferences that can be drawn from each model are abbreviated at the bottom of the table

Conditional if $p$ then $q$		Bi-conditional if and only if $p$ then $q$	
Initial model	Fleshed-out model	Initial model	Fleshed-out model
[ $p$ ] $q$	[ $p$ ] $q$ ...                not $p$ $q$ not $p$ not $q$	[ $p$ ]      [ $q$ ] ...                not $p$	[ $p$ ]      [ $q$ ] not $p$ not $q$
(a)	(b)	(c)	(d)
MP	MP MT	MP AC	MP DA AC MT

If the conditional interpretation is adopted but it is not fleshed out (Table 12.9 (a)) then people will only turn the  $A(p)$  card. This is because the mental model indicates that a  $p$  card ( $A$ ) must be paired with a  $q$  (2). So, if a card has a 2 on the other side it suggests the rule is true but if it has a 7 ( $\text{not } q$ ) on the other it falsifies the rule. The 2 card, however, will be consistent with the rule being true whatever is on its other side.

If the conditional interpretation is adopted and fleshed out ((b) in Table 12.9) then people will turn the  $A(p)$  and the 7 ( $\text{not } q$ ) cards. The reason for the selection of the  $A(p)$  card is the same as in (a). The 7 ( $\text{not } q$ ) card is now selected because it is represented as having to be paired with a  $\text{not } p$  and so if it has an  $A(p)$  on the other side it falsifies the rule.

## ACTIVITY 12.5

See if you can work out which cards should be selected for the bi-conditional interpretations and why. In each of the Tables 12.9(c) and 12.9 (d) look to see which pairs must go together in each model. A clue is given by the identities between card selections and conditional inferences in the mental logic section.

The matching effect is given exactly the same explanation in mental models theory as in mental logic. That is, it is a product of having to process implicit negations (Evans and Handley, 1999; Johnson-Laird and Byrne, 2002).

### 5.1.3 Probabilistic approach

The probabilistic approach suggests that selecting the  $p$  and the  $q$  card, far from being irrational, is in fact the *optimal* response. The general idea is quite simple, although the mathematics can be a bit off putting (Oaksford and Chater, 1994, 1996,

2003b). For the adventurous, a worked example is provided in Box 12.1 below, though don't worry if this is too off-putting – the ideas behind the calculations are explained here. The central idea in the selection task is that people are looking for the most informative evidence to help them decide whether a rule is true or false. For example, people might have to decide whether John's having a runny nose is more often associated with having a cold than one would expect by chance. If it is, then there is a predictive relationship such that his having a runny nose allows you to predict that he has a cold – they are *dependent*. If there is no such relationship, then they are *independent*.

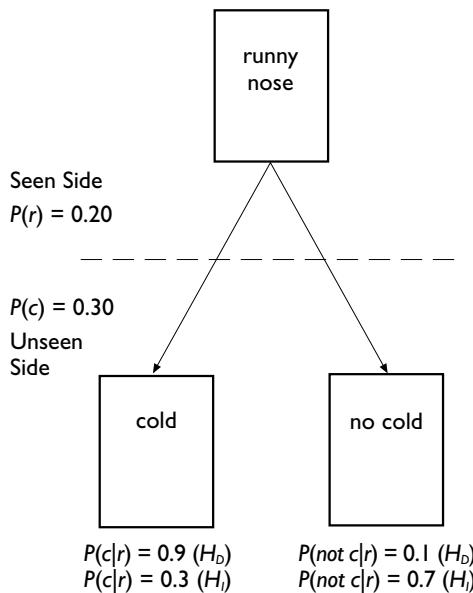
To determine which cards are more informative, information is quantified as *bits* of information. People are initially assumed to be maximally uncertain about the relationship between runny noses and colds. That is, the two possible hypotheses – that (1) they are dependent and (2) that they independent – are given an even or equal chance (0.5) of being true or false. This means that people's uncertainty is the highest it can be at 1 bit. Turning cards to reveal data can reduce this uncertainty. So, for example, turning the *runny nose* (*p*) card to find that *John did not have a cold (not q)* on this occasion, should reduce my uncertainty about which hypothesis is true. I should now feel more certain that for John runny noses and colds are independent (see Box 12.1). This reduction in my uncertainty is called **information gain**. However, participants in the selection task don't actually turn the cards over – they merely state which they would turn over! So what gets calculated is *expected* information gain. This is the reduction in uncertainty averaged over the two possibilities, that is, the other side of the *runny nose* card could reveal that John had a cold or that he did not, on this occasion.

Expected information gain can be calculated for each card (Box 12.1 shows the calculation for just the *p* card [*runny nose*]). Calculating the relevant probabilities of what is on the other side for each card involves the same three probabilities as in the conditional inference task. That is, the probability that John has a runny nose,  $P(r)$ , the probability that he has a cold,  $P(c)$ , and the probability that he has a cold given he has a runny nose,  $P(c|r)$ . If  $P(r)$  and  $P(c)$  are both low (as in Box 12.1, that is, 0.2 and 0.3 respectively), which is called the **rarity assumption**, then the expected information gain for the *q* card (*John has a cold*) is higher than for the *not q* card (*John does not have a cold*). Therefore, this model explains the standard finding in the abstract task as a *rational* consequence of trying to identify the most informative data.

We can perhaps see why this happens intuitively using an alternative example of a rule: *if a pan drops in the kitchen, then it makes a clang noise*. According to logic, testing this hypothesis exhaustively (i.e. in the real world, rather than the restricted circumstances of the selection task) would involve investigating every instance of not hearing a clang noise, to see whether a pan has dropped noiselessly. This clearly makes no sense because hearing a clang noise is a very rare event.

## 12.1

## Calculating expected information gain



$H_D$  Dependence Hypothesis: if runny nose then cold  $P(c|r) = 0.9$

$H_I$  Independence Hypothesis:  
No relationship  $P(c|r) = P(c) = 0.3$

Assume *runny nose* card is turned to reveal *no cold*, this is the data,  $D$ .

Belief before turning the card :  
 $P(H_D) = P(H_I) = 0.5$

Uncertainty Before:  $I(H) = \sum_i P(H_i) \log_2 \left( \frac{1}{P(H_i)} \right) = 1 \text{ bit}$

Uncertainty After:  $I(H|D) = \sum_i P(H_i|D) \log_2 \left( \frac{1}{P(H_i|D)} \right)$

$$P(H_D|D) = \frac{P(D|H_D)P(H_D)}{\sum_i P(D|H_i)P(H_i)} = \frac{0.1 \times 0.5}{0.1 \times 0.5 + 0.7 \times 0.5} = 0.125, \text{ and so } P(H_I|D) = 0.875,$$

i.e. highly probable  $H_I$  is true.

Therefore, uncertainty after =  $0.125 \log_2 (1/0.125) + 0.875 \log_2 (1/0.875) = 0.544$  bits

Amount by which uncertainty has reduced, or *Information Gain* =  $I(H_I) - I(H_D|D) = 1 - 0.544 = 0.456$  bits. However, the card is never turned. So the expected uncertainty after turning the card is calculated. By similar means we can calculate the uncertainty after finding *cold* = 0.811. Expected uncertainty =  $0.5 \times (0.9 + 0.3) \times 0.811 + 0.5 \times (0.1 + 0.7) \times 0.544 = 0.704$ .

Therefore, the expected reduction in uncertainty, or *expected information gain*, after turning the *runny nose* card =  $1 - 0.704 = 0.296$  bits.

This probabilistic account also explains the matching effect as a rational consequence of manipulating probabilities (Oaksford and Chater, 1994, 2003b). The probability that John does not have a cold is greater than the probability that he does: most people are cold free most of the time. So, apart from being implausible, a rule like *if John has a runny nose, he does not have a cold*, introduces a high probability event in the consequent. It turns out that if either the antecedent or consequent has a high probability, then the expected information gain for the *not q* card is higher than for the *q* card. So people should select the *John has a cold* (the *not q*) card for this rule as well, which is the matching effect.

## 5.2 The deontic selection task

Deontic reasoning is reasoning about what you should or should not do. Research in this area has recently engendered the most controversy in reasoning research. This is largely because of the strong claim made by evolutionary psychology that this research reveals the effects of various innate cognitive modules. In this section we will therefore also consider this approach to human reasoning.

Some early work on the selection task seemed to show that logic-like performance was observed when *real-world* materials, as opposed to abstract materials, were used (Wason and Shapiro, 1971). People seemed to select the more logical  $p$  and  $\text{not } q$  cards for rules like (12) and (13) below:

12     If Johnny travels to Manchester, then he takes the train.

13     If you use a second class stamp, then you must leave the envelope unsealed.

However, it was found that rules like (12) only sometimes led to logical responses whereas rules like (13) (and [14] below), for which participants were told to imagine that they were immigration officials at the airport and the cards represented immigration forms, reliably produced logic-like performance (Cheng and Holyoak, 1985).

14     If you are entering the country, then you must have a cholera inoculation.

(12), and all of the conditionals we have considered prior to this section, are called **indicative conditionals** – they describe the world or how someone behaves in it. (13) and (14) are **deontic conditionals** – they are prescriptive and state how people *should* or *should not* behave. This marks an important distinction (Manktelow and Over, 1987) for only deontic conditionals appear to produce logic-like performance reliably. Now logic indicates how the truth of a conditional depends on the truth of its antecedent and consequent, but note that the selection task for a deontic conditional cannot be solved using standard logic. Whereas finding out that Johnny travelled to Manchester by car brings into question the truth of (12), finding out someone entered the country without a cholera inoculation does *not* question the *truth* of (14). (14) can remain in force regardless of the number of people found violating it. So, for deontic conditionals, correctly selecting the  $p$  and  $\text{not } q$  cards now has nothing to do with the *truth* of the rule, and so nothing to do with logic.

This point came sharply in to focus when researchers began to investigate the factors that affect reasoning with deontic conditionals. The rule illustrated by Example 14 is an *obligation* rule, that is, it describes the pre-conditions (having an inoculation against cholera) that you are *obliged* to satisfy to carry out an action, that is, entering the country. These deontic rules can also be framed as **permission rules** (Cosmides, 1989) as in (14').

14' If you have a cholera inoculation, then you may enter the country.

This rule now describes an action you are *permitted* to carry out if you satisfy the pre-condition of having a cholera inoculation.

There are two potentially important differences between Examples (14') and (14). First, in (14') the pre-condition is now in the antecedent of the rule and the action is in the consequent. Second, in (14') *must* has changed to *may*, reflecting the fact that one is allowed, but not obliged, to enter the country on having a cholera inoculation. The first difference has an important implication. An immigration official needs to identify people who break immigration laws – people trying to enter the country without a cholera inoculation – regardless of whether the rule is expressed as in (14') or (14). For (14), as we have seen, this corresponds to selecting the *p* and *not q* cards. However, in (14') these antecedents and consequents are swapped around. Now, entering the country becomes the logical consequent *q*, and having a cholera inoculation the logical antecedent *p*. So, if participants are looking for potential law breakers, as they are asked to do in this task, then they should select the *not p* and *q* cards. Cosmides (1989) observed exactly this behaviour. This pattern of responses would appear to have nothing to do with the logic of the conditional. Other manipulations have revealed similar performance (Manktelow and Over, 1991).

In sum, the deontic selection task seems to reveal the conditions under which people can reason logically. However, the factors that affect deontic reasoning, that is, the nature of the rules, *obligation* vs. *permission*, show that it is not *logical* reasoning that is facilitated. This is because the *not p* and *q* card selection cannot be predicted from logic, although it makes perfect sense for the rules introduced in the deontic selection task. We now look at how these findings have been interpreted by the different psychological theories of reasoning.

### 5.2.1 Mental logic

Mental logicians have not explicitly addressed the deontic selection task. However, philosophers and mathematicians have formulated logical theories to account for the meanings of words such as *must* and *may*, which feature in deontic conditionals. So, in principle, mental logic might be extended to account for these inferences in the future (Manktelow and Over, 1987).

### 5.2.2 Mental models

In its basic principles the mental models account of the deontic selection task is continuous with the explanation provided for the standard selection task and indeed for all reasoning. People represent possibilities and identify counter-examples. The crucial distinction for explaining the deontic selection task is that what is deontically possible or permissible does not correspond to what is factually possible (Johnson-Laird and Byrne, 2002). So for example, if it is true that *if I turn the key then the car starts* then the three mental models as in Table 12.9 (b) would represent factual possibilities (the model in which the key is turned but the car does not start is not possible).

However, for (14), all four truth table cases are factually possible – even the case where someone enters the country without a cholera inoculation. The deontic rule does not say that this does *not* happen, it says it *should* not happen, that is, it is factually possible but deontically impermissible. This means that the representation of (14) is as follows (see Figure 12.3):

Factual possibilities		Deontic possibilities	
factually possible	<i>Entering</i>	<i>Cholera inoculation</i>	deontically possible
factually possible	<i>Not entering</i>	<i>Cholera inoculation</i>	deontically possible
factually possible	<i>Not entering</i>	<i>No cholera inoculation</i>	deontically possible
factually possible	<i>Entering</i>	<i>No cholera inoculation</i>	deontically impossible

**Figure 12.3** Factual and deontic possibilities represented in a fleshed-out mental model for the rule *if you are entering the country, then you must have a cholera inoculation*

That is, people explicitly label or mentally tag the different possibilities indicating whether they are permissible (deontically possible) or impermissible (deontically impossible). Exactly the same mental representation is formed of Example (14'), *if you have a cholera inoculation, then you may enter the country*. So, in both cases people are looking for cases that are deontically impossible. This explains why people select logically different cases – for (14) they select *p* and *not q* and for (14') they select *not p* and *q*. Even though they are logically different, these choices represent the same deontic case (Johnson-Laird and Byrne, 2002).

### 5.2.3 The probabilistic approach

The probabilistic approach to the deontic selection task (Oaksford and Chater, 1994) adopts a decision-theoretic framework first proposed by Manktelow and Over (1987, 1991). In decision theory (as you saw in Chapter 11) people are deemed to make choices that help to maximize *expected utility*, where utilities are the values people place on various outcomes. In the deontic selection task, the instructions ask people to place a high value on instances of unfairness – for instance, where someone enters the country without having had a cholera inoculation.

Suppose you are to enforce (14). Your goal is to find people entering the country who do not have a cholera inoculation. You might assign a high positive utility to this case; but the remaining possibilities are uninteresting and so are assigned a small (possibly negative) utility. Suppose also that you have no prior knowledge of the likelihood of people having had a cholera inoculation or of them trying to enter the country (so all possibilities are equally likely). Then Oaksford and Chater's (1994) formal model can be illustrated by annotating utilities and probabilities to a mental model (Johnson-Laird *et al.*, 1999) (see Table 12.10).

**Table 12.10** A mental model for the rule if you are entering the country, then you must have a cholera inoculation annotated with utilities and probabilities for each case

		Utilities	Probabilities
Entering	Cholera inoculation	-0.1	0.25
Not entering	Cholera inoculation	-0.1	0.25
Not entering	No cholera inoculation	-0.1	0.25
Entering	No cholera inoculation	5	0.25

Given these assumptions it turns out that the cards with greatest expected utilities are the *entering* (*p*) and *no cholera inoculation* (*not q*) cards and so, according to the principle of maximizing expected utility, you should therefore pick the *p* and *not q* cards. (Box 12.2 shows how these expected utilities are calculated.)

## 12.2

## Methods

### Calculating expected utilities

Given the utilities and probabilities in Table 12.10, it is possible to calculate expected utility associated with turning each card. For example, for the card marked *entering* we need to consider two probabilities: the probability that this person has had a cholera inoculation and the probability that they have not. The first, the probability of the person having had an inoculation given they are trying to enter the country is 0.5, that is, the probability of entering with an inoculation (0.25) divided by the probability of entering the country (0.5). A similar calculation gives the same value (0.5) for the second probability, that the person does not have an inoculation given they are trying to enter. These probabilities (*P*) are then multiplied by the corresponding utilities (*U*) and summed to provide the expected utility (EU) associated with turning the card. So:

$$\begin{aligned} EU(\text{entering}) &= P(\text{cholera inoculation} \mid \text{entering}) \times U(\text{entering, cholera inoculation}) \\ &\quad + P(\text{no cholera inoculation} \mid \text{entering}) \times U(\text{entering, no cholera inoculation}) \end{aligned}$$

and therefore:

$$EU(\text{entering}) = 0.5 \times -0.1 + 0.5 \times 5 = 2.45$$

Similar calculations can be carried out for each card:

$$EU(\text{not entering}) = 0.5 \times -0.1 + 0.5 \times -0.1 = -0.1$$

$$EU(\text{cholera inoculation}) = 0.5 \times -0.1 + 0.5 \times -0.1 = -0.1$$

$$EU(\text{no cholera inoculation}) = 0.5 \times 5 + 0.5 \times -0.1 = 2.45$$

This decision-theoretic account makes the same predictions as mental models theory. However, it also suggests that people's deontic reasoning should be sensitive to manipulations of utility and probability and there is evidence that seems to support this suggestion (Kirby, 1994; Manktelow *et al.*, 1995).

### 5.2.4 Evolutionary psychology

Evolutionary psychology sees many cognitive mechanisms as innately specified, having adapted under evolutionary pressures to cope with problems confronted by

early humans. Evolutionary psychologists have also argued that deontic reasoning might be under the control of innately specified cognitive modules (Cosmides, 1989; Fiddick *et al.*, 2000).

Many of the effects observed in the deontic selection task can be explained by assuming that there is a cognitive module for social contracts that govern the operation of social exchanges (Cosmides, 1989; Fiddick *et al.*, 2000). A social exchange involves satisfying a requirement in order to receive a benefit from another individual or a group. This can be expressed as a social contract, either in the form of an obligation rule (15) or a permission rule (16).

15     *If you accept the benefit then you must satisfy the requirement*

16     *If you satisfy the requirement then you are entitled to the benefit.*

The cognitive module specialized for reasoning about social contracts would be largely insensitive to the logic of the conditionals used to describe them. What is important for survival out on the savannah is not whether the rule is *true* but whether you get *cheated*. Consequently, you should look out for people who take the benefit but do not fulfil the requirement. For (14), this corresponds to someone entering the country without having had a cholera inoculation. Consequently, this account can explain the results on the standard deontic selection task.

What distinguishes this account from other explanations? The point of invoking cognitive modules is that their processing is automatic and will tend to override any domain general reasoning processes. Moreover, they are domain specific, and so (15) and (16) should only apply to situations where there is a clear *benefit–requirement* relationship. Cosmides (1989) constructed two task versions using rules like (17).

17     *If a student is to be assigned to Grover High School, then that student must live in Grover city.*

In one version of the task participants were told that going to Grover High (the *p* case) was a benefit compared to going to Hanover High (the *not p* case). In another version, this information was not included, so although the obligation to live in Grover city was stated, there was no suggestion that going to Grover High was a benefit. Far more people selected the *p* and *not q* cards when the benefit was mentioned explicitly than when it was not. Consequently, it would appear that the obligation rule form is not sufficient to produce the *p* and *not q* response – the *p* and *q* cases must be understood as benefit and requirement respectively.

Further experiments appeared to show that people have an automatic understanding of social exchange situations in the absence of any explicit rules (Fiddick *et al.*, 2000). In one condition, participants were given the rule *if you give me some potatoes, then I will give you some corn*. In another condition, participants were told to imagine they were a farmer who walks into the neighbouring village and meets someone who says *I want some potatoes* to which they respond *I want some corn*. Participants are then given four cards corresponding to four people marked:

you gave this person potatoes, you gave this person nothing, this person gave you corn, and this person gave you nothing. Participants are asked to check whether any of the people represented by the cards have cheated them. Both groups performed equally well. Now this could be because people interpret the rule-less version as involving the rule. However, Fiddick *et al.* (2000) observed that people could translate the rule-less scenario into any one of four different underlying rules, any one of which would be consistent with a social exchange, but only one of which could produce the observed results. Consequently, it would seem that reliable deontic selection task performance requires the appropriate *benefits* and *requirements* to be specified but is independent of the use of a conditional rule. So, an explanation of these tasks does not seem to involve the logic of the conditional. (However, this interpretation has been the subject of intense debate, e.g. Sperber and Girotto, 2002.)

## Summary of Section 5

---

- There are two major reasoning paradigms that use the Wason selection task.
  - The standard abstract task and the matching effect.
  - The deontic selection task.
- *Mental logic* suggests that people use the card face they can see to draw conditional inferences, though it has not been extended to the deontic task.
- *Mental models* suggest that people check their mental models for cases relevant to the abstract rule and, for deontic cases, tag the factual possibilities according to their deontic possibility.
- *The probabilistic approach* suggests that people select cards that carry most information about the relationship expressed in the conditional and, in the deontic task, select cards that have greatest expected utility.

## 6 Conclusion

We have seen how the principal theories of reasoning account for the most researched experimental tasks. The ability to explain these results is one main criterion by which to judge these theories. As we saw, they all fared reasonably well. In this final section, we continue evaluating these different theories, in particular for what they have to say about the issue of human rationality. We also take the opportunity to introduce some further evidence that might decide between these theories. In evaluating theories, there are two possible approaches we might consider, **competitive** and **integrative**.

In a competitive approach, each theory is regarded as in competition to be the one true theory of reasoning. The idea is that the proponents of each theory fight their own corner, attempting to find the killer argument or experiment that will support their theory and falsify all the others. This approach tends to lead to acrimonious exchanges in the literature. However, rarely is any argument or evidence regarded as

fatal. Indeed, in this and in many other areas of psychology, such wrangles usually end up in some kind of compromise position where it is conceded that each theory probably has its own merits and proper domain of application. Consequently, the final position arrived at is often an *integration* of theoretical positions. In this light, we first look at the relative merits of each theory before closing this chapter by looking to integrative approaches.

## 6.1 Theoretical evaluation

In this section we look at each theory and examine further evidence to distinguish between these theories where it exists. However, this evaluation falls short of plumping for one theory over another.

### 6.1.1 Mental logic

Mental logic theories have several advantages:

- They are formally very well specified, so theoreticians can prove mathematically what these theories predict.
- They preserve a full logical conception of what it is to reason rationally.

However, they also have some disadvantages:

- It is unclear how mental logic can apply to a range of data (for example, the graded phenomena observed in the suppression experiments). Consequently, the range of coverage of the theory is quite narrow.

Recently, this has led some mental logicians (Rips, 2001, 2002a and b) to propose a sharp distinction between *non-deductive* and *deductive* reasoning. The former are evaluated in terms of how probable the premises make the conclusion, which Rips calls *inductive strength*. People may be able to evaluate arguments for both *inductive strength* and *deductive validity*. This last issue points to a possible integration whereby mental logic deals with clear-cut cases of deductive reasoning and other theories, perhaps the probabilistic approach, deal with the rest.

### 6.1.2 Mental models

Features in favour of mental models include:

- The range of coverage of the data. For many phenomena in human reasoning mental models provide the only existing account. While mental logic may have the advantage of depth, mental models has the advantage of breadth.
- More reasoning researchers work in this framework than in any other.

However, as with mental logic there are problems:

- Pragmatic modulation does not seem consistent with graded effects. If people take a conditional to be true then, logically speaking, there can be no need to search for counter-examples.

There are also some more general issues:

- The attempt to provide the ‘crucial experiment’ that clearly falsifies mental logic while supporting mental models has largely been unsuccessful (though some

recent work could be argued to play this role, e.g. Johnson-Laird and Savary, 1999).

It is difficult to gauge what mental models theory says about human rationality. Although initially motivated by the logical meanings associated with the structure-building words, mental models theory has been extended well beyond the scope of standard logic. Within the scope of standard logical inference, mental models can be seen as preserving human rationality because it *approximates* logical reasoning. However, beyond the scope of standard logical inference, mental models theorists rarely show that their theories approximate any logical or mathematical theory of reasoning. Consequently, in these domains it is difficult to tell whether the theory preserves human rationality or not. This is where the mental logic theory and the probabilistic approach agree – both attempt to preserve human rationality by showing that most reasoning behaviour approximates to either logic or probability theory.

The cognitive neuroscience of reasoning may also address the issue of whether people reason with a language-based mental logic or more imagery-based mental models. For example, in a recent neuro-imaging study of people performing conditional inference and other reasoning tasks (Goel *et al.*, 1998), it was found that activation was primarily restricted to the left hemisphere language centres, rather than the right hemisphere imagery systems. Results like this seem to argue for a mental logic approach. However, such data are far from conclusive but it is certainly an interesting future direction for reasoning research.

### 6.1.3 Probabilistic approach

The general advantages of the probabilistic approach are:

- Much more of human reasoning behaviour can be seen as rational but account must be taken of people's prior knowledge of the environment, for example, the rarity assumption.
- Predictions can be derived for how manipulating probabilities and utilities should affect reasoning performance and these have generally been confirmed.

The disadvantages of the probabilistic approach are:

- The coverage of the probabilistic approach is small compared to the mental models approach.
- The theory only provides an account of how the cognitive system should behave given certain inputs. It does not provide an account of the cognitive representations and processes involved that some feel is the proper level of psychological explanation.

It is important to bear in mind that this theory suggests that the standard of rationality should change. Rather than judge human reasoning by logical standards, it should be judged by a probabilistic standard. When it is, a lot more of people's behaviour can be viewed as rational than the early experiments on human reasoning led us to expect.

#### 6.1.4 Evolutionary psychology

It is difficult to judge the evolutionary psychology approach by the same standards as the other theories because even compared to the mental logic and probabilistic approaches its scope is extremely limited. However, within the domain of deontic reasoning there is some further neuro-psychological evidence that may be relevant. The evolutionary approach suggests that people have two innate cognitive modules, one for social contracts and one for reasoning about *hazard management*. The latter involves reasoning about rules like, *if you clear up blood, you must wear rubber gloves* (Manktelow and Over, 1991). The rule indicates the precautions you should take if you encounter a hazardous situation. According to the domain general theories, such rules are dealt with by the same mechanisms that deal with social contract rules. However, there is recent evidence of a neuro-psychological patient with brain damage, who shows an impaired ability to reason about social contracts but an intact ability to reason about hazard management rules (Stone *et al.*, 2002). This seems to suggest that this patient has an intact innate hazard management module but a damaged innate social contract module. However, unless a patient is found with the opposite deficit, that is, impaired hazard management reasoning and intact social contract reasoning, these results remain inconclusive.

### 6.2 Integration, dual processes and individual differences

Attempts to integrate theories of reasoning centre on dual process theories that have a long pedigree in reasoning research (Evans, 1984; Evans and Over, 1996; Stanovich and West, 2000). These theories suggest a two-way partition in reasoning abilities. As we have already suggested, this is similar to some mental logic theorists who have invoked the distinction between deductive and non-deductive reasoning (Rips, 2002b). Typically, these theories suggest that we do have a, perhaps limited, ability for explicit logical reasoning that may be embodied in mental logic or in mental models. However, a lot of reasoning goes on implicitly and is independent of these logical processes. More recently the distinction has been drawn between two types of rationality (Evans and Over, 1996). People are rational in one sense when their reasoning conforms to a normative standard like logic. They are rational in another sense when they reason in order to achieve their goals in the world, regardless of whether their reasoning conforms to a normative standard. Different mental processes are involved in these forms of reasoning.

One recent source of evidence for this approach is the study of individual differences (Stanovich and West, 2000). For example, it has been shown that the ability to make the logically correct response on the selection task is associated with IQ (Stanovich and West, 1998). It would appear that participants with a high IQ are capable of interpreting this task logically (and choosing the *p* and *not q* cards). However, when you consider just the remaining participants, then IQ seems to correlate with the non-logical but standard *p* and *q* card response (Newstead *et al.*, 2004). This evidence seems to argue for a dual-process theory. Perhaps people possess automatic unconscious reasoning mechanisms that operate in accordance with probabilistic standards of reasoning (explaining the *p* and *q* cards' selection). However, those with higher IQs may be capable of ignoring the prior knowledge that

is required to determine the relevant probabilities, and can then reason logically about the task.

Such integrative approaches are also consistent with the trend among some mental model theorists to add probabilistic components to the core theory (Schroyens and Schaeken, 2003). The critical question then becomes the balance of reasoning processes. That high IQ is apparently associated with logical responses suggests that perhaps most human reasoning is carried out by unconscious, probabilistic or inductive processes. However, the jury is still very far from delivering a verdict on this question. Nonetheless, the emergence of integrative approaches should be seen as a positive sign. This is because it opens up the area of human reasoning to more interesting possibilities, other than that my theory is right and yours is wrong!

## **Answer to Activity 12.2**

Possibility C shows AC to be an invalid form of inference. In this, both *if p then q* and *q* are true, but *p* is false. Possibility C also shows DA to be invalid: though *if p then q* and *not p* are true, *not q* is false.

## **Further reading**

- Braine, M.D.S. and O'Brien, D.P. (eds) (1998) *Mental Logic*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Johnson-Laird, P.N. and Byrne, R.M.J. (1991) *Deduction*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Oaksford, M. and Chater, N. (1998) *Rationality in an Uncertain World*, Hove, Psychology Press.
- Manktelow, K.I. (1999). *Reasoning and Thinking*, Hove, Psychology Press.

## **References**

- Boole, G. (1854) *An Investigation of the Laws of Thought On Which are Founded the Mathematical Theories of Logic and Probabilities*, Cambridge, Macmillan and Co.
- Braine, M.D.S. and O'Brien, D.P. (eds) (1998) *Mental Logic*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Byrne, R.M.J. (1989) ‘Suppressing valid inferences with conditionals’, *Cognition*, vol.31, pp.1–21.
- Byrne, R.M.J., Espino, O. and Santamaria, C. (1999) ‘Counter-examples and the suppression of inferences’, *Journal of Memory and Language*, vol.40, pp.347–73.
- Cheng, P.W. and Holyoak, K.J. (1985) ‘Pragmatic reasoning schemas’, *Cognitive Psychology*, vol.17, pp.391–416.
- Cosmides, L. (1989) ‘The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task’, *Cognition*, vol.31, pp.187–276.

- Cummins, D.D., Lubart, T., Alksnis, O. and Rist, R. (1991) 'Conditional reasoning and causation', *Memory and Cognition*, vol.19, pp.274–82.
- Cummins, D.D. (1995) 'Naive theories and causal deduction', *Memory and Cognition*, vol.23, no.5, pp.646–58.
- Evans, J. St B.T. (1977) 'Linguistic factors in reasoning', *Quarterly Journal of Experimental Psychology*, vol.29, pp.297–306.
- Evans, J. St B.T. (1984) 'Heuristic and analytic processes in reasoning', *British Journal of Psychology*, vol.75, pp.451–68.
- Evans, J. St B.T., Clibbens, J. and Rood, B. (1996) 'The role of implicit and explicit negation in conditional reasoning bias', *Journal of Memory and Language*, vol.35, no.3, pp.392–409.
- Evans, J. St B.T. and Handley, S.J. (1999) 'The role of negation in conditional inference', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.52A, no.3, pp.739–69.
- Evans, J. St B.T. and Lynch, J.S. (1973) 'Matching bias in the selection task', *British Journal of Psychology*, vol.64, pp.391–7.
- Evans, J.B. St. B.T. and Over, D.E. (1996) *Rationality and Reasoning*, Hove, Psychology Press.
- Evans, J.B. St B.T., Handley, S.J. and Over, P.E. (2003) 'Conditionals and conditional probability', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.29, pp.321–35.
- Fiddick, L., Cosmides, L. and Tooby, J. (2000) 'No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task', *Cognition*, vol.77, no.1, pp.1–79.
- Goel, V., Gold, B., Kapur, S. and Houle, S. (1998) 'Neuroanatomical correlates of human reasoning', *Neuropsychologia*, vol.29, pp.901–9.
- Inhelder, B. and Piaget, J. (1958) *The Growth of Logical Reasoning*, New York, Basic Books.
- Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S. and Caverni, J.P. (1999) 'Naive probability: a mental model theory of extensional reasoning', *Psychological Review*, vol.106, no.1, pp.62–88.
- Johnson-Laird, P.N. and Savary, F. (1999) 'Illusory inferences: a novel class of erroneous deductions', *Cognition*, vol.71, no.3, pp.191–229.
- Johnson-Laird, P.N. and Wason, P.C. (1970) 'Insight into a logical relation', *Quarterly Journal of Experimental Psychology*, vol.22, no.1, pp.49–61.
- Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge, Cambridge University Press.
- Johnson-Laird, P.N. and Byrne, R.M.J. (eds) (1991) *Deduction*, Hillsdale, NJ, Erlbaum.
- Johnson-Laird, P.N. and Byrne, R.M.J. (2002) 'Conditionals: A theory of meaning, pragmatics, and inference', *Psychological Review*, vol.109, no.4, pp.646–78.
- Kirby, K.N. (1994) 'Probabilities and utilities of fictional outcomes in Wason's four card selection task', *Cognition*, vol.51, pp.1–28.

- Manktelow, K.I. and Over, D.E. (1987) 'Reasoning and rationality', *Mind and Language*, vol.2, pp.199–219.
- Manktelow, K.I. and Over, D.E. (1991) 'Social roles and utilities in reasoning with deontic conditionals', *Cognition*, vol.39, pp.85–105.
- Manktelow, K.I., Sutherland, E.J. and Over, D.E. (1995) 'Probabilistic factors in deontic reasoning', *Thinking and Reasoning*, vol.1, pp.201–20.
- Newstead, S.E., Handley, S.J., Harley, C., Wright, H. and Farrelly, D. (2004) 'Individual difference in deductive reasoning', *Quarterly Journal of Experimental Psychology*, vol.57, pp.33–60.
- Oaksford, M. and Chater, N. (1994) 'A rational analysis of the selection task as optimal data selection', *Psychological Review*, vol.101, pp.608–31.
- Oaksford, M. and Chater, N. (1996) 'Rational explanation of the selection task', *Psychological Review*, vol.103, pp.381–91.
- Oaksford, M. and Chater, N. (1998) *Rationality in an Uncertain World: Essays on the Cognitive Science of Human Reasoning*, Hove, Psychology Press.
- Oaksford, M. and Chater, N. (2003a) 'Computational levels and conditional inference: Reply to Schroyens and Schaeken', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.29, no.1, pp.150–6.
- Oaksford, M. and Chater, N. (2003b) 'Optimal data selection: revision, review and re-evaluation', *Psychonomic Bulletin and Review*, vol.10, pp.289–318.
- Oaksford, M. and Chater, N. (2003c) 'Probabilities and pragmatics in conditional inference: Suppression and order effects' in Hardman, D. and Macchi, L. (eds) *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*, London, John Wiley and Sons, pp.95–122.
- Oaksford, M., Chater, N. and Larkin, J. (2000) 'Probabilities and polarity biases in conditional inference', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.26, no.4, pp.883–99.
- Quinn, S. and Markovits, H. (2002) 'Conditional reasoning with causal premises: evidence for a retrieval model', *Thinking and Reasoning*, vol.8, no.3, pp.179–91.
- Rips, L.J. (1994) *The Psychology of Proof*, Cambridge, MA, MIT Press.
- Rips, L.J. (2001) 'Two kinds of reasoning', *Psychological Science*, vol.121, no.2, pp.129–34.
- Rips, L.J. (2002a) 'Reasoning imperialism' in Elio, R. (ed.) *Common Sense, Reasoning and Rationality*, New York, Oxford University Press, pp.215–35.
- Rips, L.J. (2002b) 'Reasoning' in Pashler, H. and Medin, D. (eds) *Steven's Handbook of Experimental Psychology, Memory and Cognitive Processes*, 3rd edn, vol.2, New York, John Wiley and Sons, pp.363–411.
- Schroyens, W. and Schaeken, W. (2003) 'A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.29, no.1, pp.140–9.
- Sperber, D. and Girotto, V. (2002) 'Use or misuse of the selection task?: rejoinder to Fiddick, Cosmides, and Tooby', *Cognition*, vol.85, no.3, pp.277–90.

- Stanovich, K.E. and West, R.F. (1998) 'Cognitive ability and variation in selection task performance', *Thinking and Reasoning*, vol.4, no.3, pp.193–230.
- Stanovich, K.E. and West, R.F. (2000) 'Individual differences in reasoning: Implications for the rationality debate?' *Behavioral and Brain Sciences*, vol.23, no.5, pp.645–64.
- Stone, V.E., Cosmides, L., Tooby, J., Kroll, N. and Knight, R.T. (2002) 'Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage', *Proceedings of the National Academy of Sciences*, vol.99, no.17, pp.11531–6.
- Taplin, J.E. (1971) 'Reasoning with conditional sentences', *Journal of Verbal Learning and Verbal Behaviour*, no.10, pp.219–25.
- Wason, P.C. (1968) 'Reasoning about a rule', *Quarterly Journal of Experimental Psychology*, vol.20, pp.273–81.
- Wason, P.C. and Shapiro, D. (1971) 'Natural and contrived experience in a reasoning problem', *Quarterly Journal of Experimental Psychology*, vol.23, pp.63–71.



---

# PART 5

# CHALLENGES, THEMES AND ISSUES

---

Introduction

**Chapter 13 Cognition and emotion**

*Jenny Yiend and Bundy Mackintosh*

**Chapter 14 Autobiographical memory and the  
working self**

*Martin A. Conway and Emily A. Holmes*

**Chapter 15 Consciousness**

*Jackie Andrade*

**Chapter 16 Cognitive modelling and cognitive  
architectures**

*Paul Mulholland and Stuart Watt*

**Chapter 17 Theoretical issues in cognitive psychology**

*Tony Stone*

# Introduction

The first four parts of this book help address the question ‘what is cognitive psychology?’, a question that we first took up in Chapter 1. Just as someone can answer the question ‘what is a cat?’ by pointing to examples of cats, so we can gain an understanding of cognitive psychology by considering different topics in cognition. As well as introducing substantive research areas, the previous 11 chapters illustrate some of the breadth and depth of the cognitive approach, its assumptions and commitments, its methods, and its successes.

Armed with an understanding of key examples of cognitive psychology, this part of the book considers two different kinds of question. The first concerns how widely the cognitive approach may be successfully applied. Many research topics, such as emotion, the self, and consciousness, have been considered at various times to lie outside the purview of cognitive psychology. Yet cognitive psychologists have recently begun to apply their approach to these problems. The difficulties in so doing, as well as the successes that arise, are discussed in Chapters 13 to 15. What becomes clear in these chapters is that consciousness, an aspect of mind also discussed in Chapter 1, has become a theme for many chapters in this book. In Chapters 13 and 14, conscious experience plays an ever more important role until, in Chapter 15, it becomes the focus of our enquiry. The second type of question concerns the kinds of explanation that are pursued in cognitive psychology, and that have also developed as themes in this book. Chapter 16 focuses on cognitive modelling and cognitive architecture, the role that models play in theories of cognition, and criteria for their evaluation. Chapter 17, taking a more philosophical approach, discusses some of the key theoretical issues within cognitive psychology – issues that have formed the basis for ongoing debates that are helping to shape our understanding of the nature of cognition.

In Chapter 13, Jenny Yiend and Bundy Mackintosh examine the relationship between cognition and emotion, aspects of the mind that Western thought has traditionally viewed as opposed to one another. Early in their chapter, the authors distinguish three components of emotions. There is a physiological or bodily response, such as heart rate (and the sweatiness of hands!) increasing at the sight of a loved one. There are emotional behaviours, in this example perhaps smiling broadly and, if the social context allows it, touching. And, third, there are emotional feelings, a glow of affection or, possibly, an intense rapture. The last of these components, the emotional feelings, are a large part of what many people think of when reflecting on their normal conscious experience. The ‘affective tone’ of a particular experiential episode is often one of its most salient features. Indeed, Jenny Yiend and Bundy Mackintosh go on to look at evidence showing that people are better able to retain information that is congruent with their current mood than information that is not. That is, they remember happy information if it is presented to them while they are happy and gloomy information presented to them when they are gloomy. In addition, retrieval of information is best when mood at encoding matches mood at retrieval. You are better able to recall what you heard in a cheerful mood when you are again in a cheerful mood.

The relationship between cognition and emotion provides strong links between Chapters 13 and 14. Affective tone is again a prominent feature of autobiographical memories, which is the topic of Chapter 14. Martin Conway and Emily Holmes are interested in the relationship between autobiographical memory and the self, and their chapter has strong links also to the memory chapters in Part 3, with, for example, the episodic/semantic distinction of Chapter 8 making a reappearance. However, research into autobiographical memories cannot rely on the same kinds of laboratory-based, experimental methods used to investigate other types of memory – think of the difficulties of running a memory experiment where participants will be tested for recall decades later! Research in this area tends to rely on particular distinctive methods aimed at eliciting memories of actual episodes from our past, and the chapter activities invite you to take part in some of these procedures.

As with the emotional feelings discussed in Chapter 13, autobiographical memory also has a self-evident bearing on consciousness; much of the stream of consciousness features episodes of autobiographical memory or is derived from such episodes and from one's sense of self. The authors of Chapter 14 develop ideas about how autobiographical memory relates to what they call the 'working self'. This notion is modelled on the concept of working memory, which was the topic of Chapter 9, and which, because it refers to a person's current cognitive activity (what they have in mind) is itself about consciousness. The working self is conceived as a hierarchy of interconnected goals, some but not all of which can enter consciousness. The chapter also examines the disruptive effects of affectively charged trauma memories that, in conditions such as post-traumatic stress syndrome, frequently intrude upon consciousness.

Consciousness has been a pervasive presence in many of the chapters in earlier parts, such as those on attention (why you are conscious of this rather than that), perception of objects and words (how you become conscious of this object or this word), recognition and categorization (what you consciously experience this object *as* or that piece of discourse *as* meaning). And, as just indicated, the affective tone of experience is also a prominent feature of Chapters 13 and 14. In all these chapters, however, consciousness has played a supporting role whilst some other aspect of cognition has played the lead. In Chapter 15, however, consciousness comes out of the shadows to take centre stage. Jackie Andrade goes head-on with this most difficult and elusive of concepts. She outlines philosophical approaches to the vexed issue of consciousness and offers an analysis of the place of consciousness within cognitive psychology. She then goes on to examine a range of empirical cognitive research into the nature of consciousness. She starts by revisiting the topic of implicit memory that was discussed in Chapter 8, and goes on to extend this into a discussion of implicit learning. The evidence for learning without conscious awareness is difficult to interpret conclusively. Certainly, some implicit learning seems possible, but as far as the modern dream of effortless sleep learning is concerned the news is entirely negative. Jackie Andrade looks into controlled versus automatic processing of information and then considers what neuropsychological evidence can tell us about the possible modularity of consciousness. Following this, she presents an analysis of what the function of consciousness might be, and this links back to ideas rehearsed in Chapter 13.

Chapters 13 to 15 thus present the results of applying the cognitive approach to difficult research questions that at one time were thought to be beyond the purview of cognitive psychology. As the authors themselves imply, cognitive psychology has succeeded in approaching these topics partly by identifying those aspects that might be amenable to the cognitive approach, and analysing them separately. It follows that there are aspects to emotion, the self and consciousness – for example, emotional *feelings*, aspects of *identity*, *phenomenal* consciousness – that as yet remain recalcitrant, and may, for all we know, continue to elude our best attempts at explanation. Nevertheless, as the chapters indicate, the application of the cognitive approach has yielded some significant insight and numerous positive results, and interest in these research areas is likely only to increase.

As mentioned at the outset of this introduction, Chapters 16 and 17 then address a somewhat different set of questions. For example, both chapters focus upon the relationship between computation and cognition, picking up on a theme first introduced in Chapter 1. Paul Mulholland and Stuart Watts's Chapter 16 has the slightly daunting title of 'Cognitive modelling and cognitive architectures'. However, the authors do a fine job of rendering this austere sounding topic comprehensible! The topic of cognitive modelling has been broached in many earlier chapters, such as Graham Hitch's Chapter 9 on working memory. In Chapter 16, the topic is illustrated and illuminated through an informative contrast between parallel distributed processing (PDP, or connectionist) models and rule-based systems. This leads the reader to the concept of a 'cognitive architecture', a concept that is needed to ensure a clear distinction between cognitive models and the computers on which they run. The notion of cognitive architecture is developed and elaborated by means of a careful overview of the ACT-R architecture. The authors go on to explain how ACT-R accounts for a range of phenomena found in the study of human memory. They then consider how ACT-R models the acquisition of arithmetic skills. Finally, they offer a comparison of rule-based and PDP architectures, followed by an analysis of the criteria against which the performance of a model should be judged.

In Chapter 17, Tony Stone provides a theoretical overview of cognitive psychology with reference to three issues, each of which has been encountered a number of times in previous chapters. He revisits the contrast between rule-based and connectionist (or PDP) models, which you will already have encountered in Chapters 1 and 16, and gives a detailed evaluation of the two types of cognitive architecture within the context of the now famous past-tense debate. He demonstrates that the argument turns upon such complex matters as what it means for a rule to be explicitly or implicitly represented, and what it means for a representation to be compositional. Tony Stone's second theoretical issue concerns the modularity, or otherwise, of the mind. The notion of cognitive modules, or systems, has cropped up repeatedly since its first mention in Chapter 1 – for example in Chapter 8 on memory and Chapter 15 on consciousness. In Chapter 17, Fodor's theory of modularity is described and evaluated within the context of the philosophy of science. Modularity turns out to be another complex topic that continues to undergo theoretical refinement. The final theoretical issue considered in Chapter 17 is the relationship between cognitive psychology and the brain. Most if not all of the

preceding chapters have included at least some appeal to neuropsychological evidence, but what is the status of such evidence? How does it bear upon cognitive theory? The discussion is introduced with reference to Marr's three levels of explanation, a further connecting loop back to Chapter 1. Tony Stone examines whether cognitive and neurobiological theories co-evolve or whether the former will eventually be reduced to the latter. Such theoretical questions as these, difficult as they are to resolve, help to provide an overarching conception of cognition, what it is, how we might model and theorize about it, and how such theories 'fit' in the wider picture of scientific enquiry.



# Cognition and emotion

## Chapter 13

Jenny Yiend and Bundy Mackintosh

---

### 1 Introduction

This chapter is concerned not just with cognition but with how emotions influence, and are influenced by, cognitive processes. Emotions are such a familiar and fundamental aspect of everyday life that it is often this very ability to experience and express emotion that is seen as a crucial distinction between the behaviour of humans and (possibly imaginary) high functioning computers or robots. As you have seen in previous chapters, it is possible for a computer to solve successfully many difficult tasks and so mimic human achievements. However, one very salient distinction between the performance of computer and human is that the computer won't show pleasure when reaching its goal nor frustration if it fails, let alone empathy with its human operator. We shall consider whether this is really an advance later in the chapter.

When emotions seem such an important part of our lives, it might come as a surprise that despite the rapid development of psychology as a discrete scientific discipline since the mid 1800s, the study of emotion has largely taken a back seat. Why should this be so? One reason is undoubtedly the behaviourist legacy. Behaviourists such as John B. Watson (1878–1958), and Burrhus Frederick Skinner (1904–1990) recognized the need for scientific rigour and objective, verifiable measurement and were therefore exclusively concerned with the overt behaviours displayed by an organism – those which could be directly observed and measured. For behaviourists, reference to unseen mental processes was taboo. Their emphasis on objectivity and empiricism continues to be an important influence in cognitive psychology today. This historical bias for a long time deterred study of emotion, in which the main component – feelings – can only be accessed through introspection. Undoubtedly, another factor has been the attitude towards emotions often expressed in Western European societies among others. At least since Plato (375 BC) emotions have been viewed as impediments to rational thought. Darwin thought of them as childish or immature responses, a residual hangover from our evolutionary past that no longer had useful functions for the mature adult. However, the study of emotion and how it interacts with cognition has enjoyed a resurgence of interest. This is largely because of the development of objective, quantifiable ways of measuring the concomitants of emotion, such as psychophysiological techniques, brain imaging, and a shift in attitude towards the importance and function of emotions in everyday life.

---

#### ACTIVITY 13.1

Stop for a moment and reflect on one recent episode when you experienced emotion and try to jot down three or more aspects that characterized this occasion as being emotional.

## COMMENT

This is often a hard task, since one common feature of emotional situations is that it is sometimes difficult to put into words what is happening! However, you may have noted your feelings and perhaps what or whom you considered was the cause of the emotion. Did you include a description of how you behaved or a change in body sensations?

One example could go something like this (note the different elements of this description):

One day when I was alone at home a special delivery van stopped at our house looking for an address nearby. I stepped outside to point out directions, the front door blew in the wind and locked behind me. 'No problem', I thought, and went to the usual hiding place to retrieve the spare key. It was missing. Now I felt anger, frustration and regret. Why wasn't the key in its usual place? I blamed others for not replacing it. Why had I been so stupid to let the door slam? I swore and banged my fist. I went back to the door and rattled the handle. I was taking deep breaths and felt my heart beating fast with annoyance. I paced rapidly up and down whilst I tried to work out what to do next ...

---

## 1.1 Components of emotion

From the example above it is clear that there are different aspects to any emotional response. Traditionally, psychologists have identified at least three characteristics that are embodied within an emotional episode. These are:

- Behaviours
- Bodily responses (physiology)
- Feelings.

### 1.1.1 Emotional behaviour and expression

Many of the behaviours associated with emotion will be familiar to you. Some simple examples include laughing when you are happy; withdrawing from something you find disgusting; becoming agitated and raising your voice when you are angry, and quiet, withdrawn and slow when you are sad. Facial expressions that are characteristic of different emotions are also examples of behavioural responses. As these are all observable phenomena, they can be easier to study empirically than internal feelings. However, they cannot generally be used to infer emotions directly since, unlike most of the bodily (physiological) responses associated with emotion, emotional expressions can be brought under some degree of control. You can suppress your smile; make a special effort to appear cheerful when sad; feign interest in order to be polite; and curb your angry behaviour if it might jeopardize your well-being. More problematic for research, various cultures and social groups differ in their code of conduct with respect to emotional expressions. For example, in some cultures, such as in many Arabian countries,

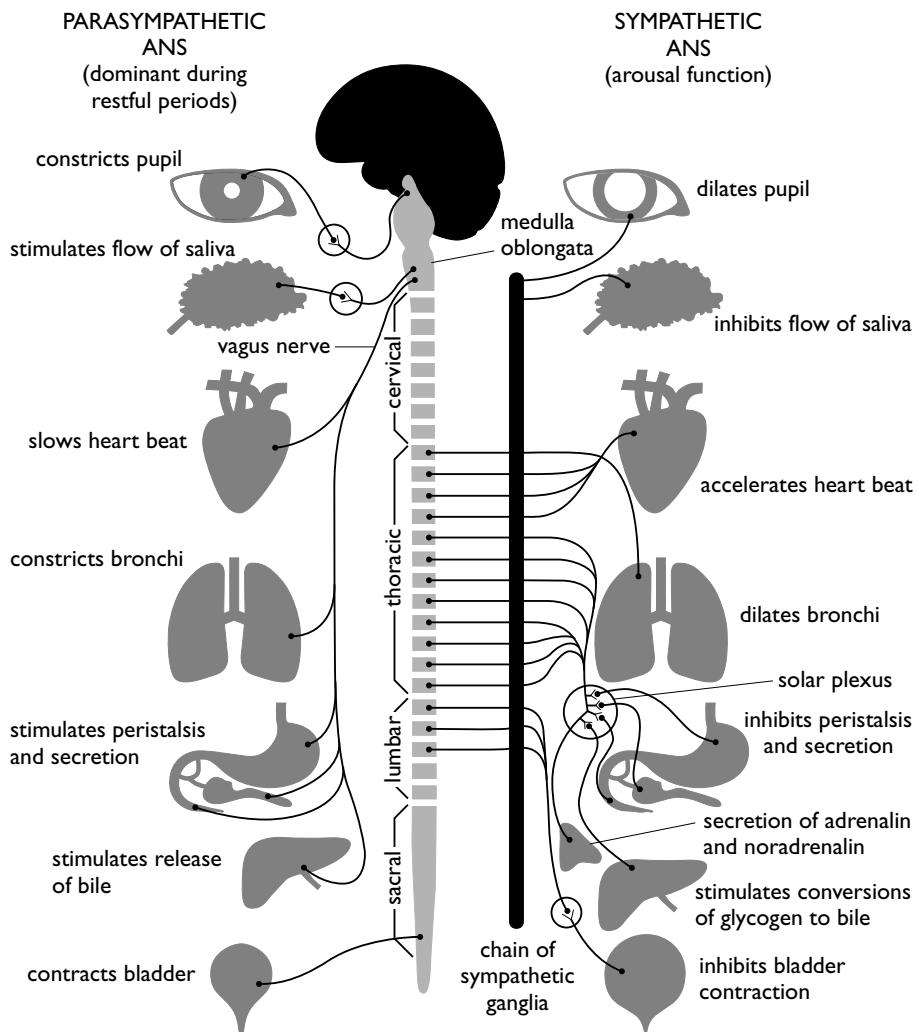
public grieving involves overt crying, moaning, and beating of the chest which are seen as appropriate expressions of respect for the dead in assembled company, whereas in others, for instance in Japan, a polite smile and tight emotional control are expected.

Unlike the very individual feeling of emotion and internal bodily response to emotion, emotional behaviour, including emotion expressions, are visible to others; that is, they can communicate (albeit imperfectly) the individual's emotional status. In any form of communication, understanding the mechanism fully requires researching both the recognition and the production of the appropriate signals (as discussed in Chapters 6 and 7). So far, much more research effort has been directed towards recognition rather than production of behavioural correlates of emotion. Facial expressions, rather than other emotion behaviour such as 'body language', have attracted most research attention.

### 1.1.2 Bodily responses

The bodily responses associated with emotions are the physiological reactions such as sweating when you feel anxious, or your heart racing when you feel agitated or excited. These reactions have been refined during evolution and are vital to survival. For example, if a lion attacks then there must be little delay before escape ('flight'), or maybe aggression ('fight'), begins. Typically an animal's body will respond to this kind of threat by diverting the blood flow away from less vital regions, such as the gut (digesting lunch suddenly is less important than ensuring you don't become someone else's). The extra blood, and therefore energy, is supplied to the major muscle blocks and the brain (rapid processing of information is needed or energy could be wasted running in the wrong direction). The contents of the blood are altered, boosting 'fuel' in the form of blood glucose, and cholesterol, and increasing clotting agent (which serves to stem blood loss in case of wounds), and so on. This physiological reaction stands us in good stead when physical action is required to ensure survival.

Many of these bodily responses are controlled by the **autonomic nervous system (ANS)** (see Figure 13.1 overleaf), a network of nerve fibres throughout the body that transmits signals to the various organs, muscles and glands. The ANS is divided into two sections. The **sympathetic ANS** produces effects associated with arousal. These include secretion of the hormone adrenalin from a gland near the kidneys. Adrenalin release initiates and enhances sympathetic activity leading to changes such as accelerating heart rate, vasoconstriction (constriction of the blood vessels), increased respiration (breathing) rate and depth, and reduced gastrointestinal (gut) activity. The intricate pattern of changes in hormone levels, breathing, redirection of blood flow and pressure and changes in its constituents, and the many other changes occurring under stress all prepare the body for physical exertion (the 'flight or fight' response described above). This is a physiological pattern fine-tuned by evolution and shared by other mammals, with only relatively minor details differing between species. In contrast, the **parasympathetic ANS** tends to dominate during periods of rest, having broadly opposing effects on the body.



**Figure 13.1** The autonomic nervous system (ANS) and physiological responses associated with emotions, showing sympathetic and parasympathetic sections

Source: Reber, 1995, p.76

Box 13.1 discusses techniques for measuring physiological responses and so for investigating this component of emotion.

### 13.1

### Methods

#### Measuring emotion using psychophysiology

Most emotional states tend to lead to increased arousal and therefore produce corresponding physiological signs. Psychologists have been able to devise ways of measuring this physiological change precisely. For example, by applying a tiny electrical current across the fingers we can measure the electrical resistance of the skin. This changes according to minute differences in the amount of sweat



produced and so provides a physiological measure known as either GSR (galvanic skin response) or SC (skin conductance), that correlates with changes in arousal (recall the use of GSR responses to detect the influence of 'shocked' words in the non-attended messages discussed in Chapter 2, Section 1.3). The measurement of changes in GSR has also been used as a 'lie detector' picking up individuals' emotional response arising during deception. Heart rate, another psychophysiological measure, is usually measured as 'beats per minute' using a simple transducer which converts the movement produced by the pulse into electrical energy. Other common measures include cortisol levels in the blood (related to adrenalin production), electromyography (EMG: muscle tension and activity, usually recorded from the face), respiration rate and surface skin temperature (related to dilation or constriction of the blood vessels).



**Figure 13.2** Measuring EMG: electrodes are shown under the eye and measure 'startle' or 'blink' magnitude, which is the muscle contraction produced when you blink or are surprised, for example in response to a sudden loud noise

### 1.1.3 Feeling emotions

Feelings are private and subjective. They are, by definition, states of experiential awareness. Around the world, humans can usually report a wide range of different states or feelings, from anger to fear to love, which can be recognized and understood by those around them. Within psychology the feeling component of emotion is inextricably bound up with notions of conscious awareness and the subjective self (see Chapters 14 and 15). Emotion researchers are often interested in whether the stimuli or tasks that they use elicit positive or negative feelings, and if so to what degree. They certainly acknowledge that feelings co-occur with the other markers of emotion, but otherwise cognitive psychologists have concentrated less on feeling states than on exploration of the cognitive processing associated with emotions and emotional information.

## ACTIVITY 13.2

---

Thinking back to what you've learned so far about cognition, would you say that emotions may also have a cognitive component as well as the components we have just discussed (behaviour, bodily responses, feelings)? If so, we would be able to study the cognitive side of emotions using the techniques and paradigms familiar to cognitive scientists.

### COMMENT

Looking only at the three components of emotion discussed above, you might well answer 'no' with regard to bodily responses and feelings. The behavioural component of emotions includes face perception, which will be familiar from Chapter 4, though in emotion research it is the emotional expression not identity that is of interest. However, as we proceed you will see that many of the tools developed by cognitive scientists can be widely adapted to the study of emotions. Not only do cognitive processes interact with emotions, but many psychologists believe that they are an integral part of producing the emotions themselves.

---

One issue that researchers face when trying to study any of the components of emotion described above is how to elicit realistic effects in the laboratory. In real life, emotions are usually stronger than those seen in the laboratory, so we have to be cautious when generalizing from the results of any given study. There are at least two reasons why it is difficult to study the behaviours, physiology and feelings associated with *strong* emotions:

- 1 It is sometimes unethical to induce strong emotions, for example strongly negative ones, in a laboratory setting.
- 2 Strong emotions are hard to elicit in a predictable fashion and take a while to die down, making laboratory study impractical.

For these reasons, you will find throughout this chapter that the study of cognition and emotion is largely confined to consideration of relatively mild emotional states and the processing of mildly emotional information. However, the information gained from such work provides useful insights about possible cognitive processing when emotion is more extreme.

# Summary of Section 1

---

- Although the study of emotions from a cognitive point of view has historically been neglected, the advent of new techniques and new ideas as to the significance and function of emotions has brought a resurgence of interest in the study of emotion and how it interacts with cognition.
- There are thought to be three main components of emotions:
  - Emotional behaviour and expression (e.g. emotional facial expressions)
  - Bodily responses (e.g. galvanic skin response)
  - Feelings.

## 2 Different emotions

You should now have some idea of what is meant by the term ‘emotion’ in psychology. Next we shall consider how one might classify and explain the huge variety of different emotions that individuals typically report. Psychologists usually take one of two approaches to dealing with the task of accounting for different emotional experiences. Some refer to a set of **basic emotions**, while others take a **dimensional** view.

### 2.1 Basic emotions

One approach has been to assume that underlying the richness of emotion experience there are a small number of discrete emotions – ones considered to be the most fundamental or important. This idea is analogous to the processing of colour by the visual system, where the whole range and subtlety of our colour experience is achieved through stimulation of just three different types of cones in the retina. Likewise, it is argued that different combinations of ‘basic emotions’ can produce all the other emotions. For example, a mixture of joy and acceptance produces friendliness according to Plutchik, a prominent basic emotions theorist.

There are several distinct challenges to the notion of basic emotions: one is to provide evidence for the existence of a small number of discrete emotion states; another is to decide how many emotions should be called basic and which ones they are. The idea of basic emotions has considerable general support but few agree exactly on the appropriate number and type of emotions that should be included. This point is illustrated by Table 13.1 (overleaf).

Despite these widely differing views there are five emotions, sometimes called **the Big Five**, that appear to represent a broad consensus among basic emotions psychologists. These are anger, fear, sadness, disgust and happiness. One of the most influential psychologists from this tradition, Paul Ekman, building on research described by Darwin in *The Expression of the Emotions in Man and Animals* (1998, first published in 1872), has collected a formidable body of information from cross-cultural studies to support the fundamental status and importance of these five emotions. He was impressed by the observation that wherever he travelled people

**Table 13.1** Basic emotion theorists and the emotions they propose

<b>Emotion theorist</b>	<b>Fundamental emotion</b>
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman, Friesen, and Ellsworth	Anger, disgust, fear, joy, sadness, surprise
Frijda	Desire, happiness, interest, surprise, wonder, sorrow
Gray	Rage and terror, anxiety, joy
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
James	Fear, grief, love, rage
McDougall	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
Mowrer	Pain, pleasure
Oatley and Johnson-Laird	Anger, disgust, anxiety, happiness, sadness
Panksepp	Expectancy, fear, rage, panic
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Watson	Fear, love, rage
Weiner and Graham	Happiness, sadness

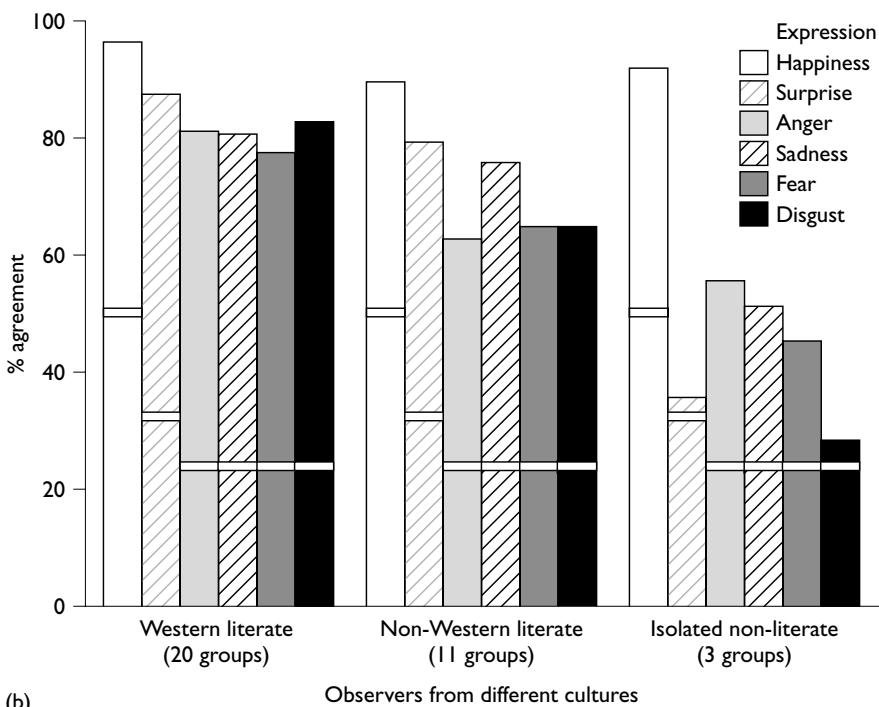
Source: Power and Dalgleish, 1997

displayed broadly similar emotions, and that he had no difficulty in interpreting them despite language barriers. For more systematic research his main method was to show pictures of facial expressions, such as those in Figure 13.3(a), and determine whether peoples from different cultures consistently select the same emotion label to describe each one.

Figure 13.3(b) shows some typical results for six emotions – the Big Five plus surprise. Although there is some variation, particularly within isolated non-literate cultures, there is always agreement above what would be expected if people were just guessing (the ‘chance’ level, shown by the white bars). Ekman also provided evidence for basic emotions in the production as well as in the recognition of expressions. He visited a visually isolated non-literate group in New Guinea (people who had not previously met or seen pictures of anyone from outside their own cultural group) and asked them to show him what their face would look like if they were sad, happy and so on. He then took videos of their expressions and played them back to American students, who had to decide which emotion was being displayed by the New Guineans. The American judges had no problem in identifying the different emotions according to the (translated) labels to which the New Guineans had been responding, which further supports the notion of pancultural or universal emotions (Ekman *et al.*, 1969; Ekman, 1999, provides a review of all his work).



(a)



**Figure 13.3** (a) Some of the photos of facial expressions used by Ekman, showing (left to right) anger, fear, disgust, surprise, happiness and sadness; (b) Results from cross-cultural studies showing differences in recognizing facial expressions of six emotions

Source: (a) Ekman and Friesen, 2003; (b) Rosenzweig et al., 1999, Figure 15.3, p.414

In support of the basic emotions approach, Ekman provides extensive evidence from cross-cultural work such as ratings of spontaneous displays of emotion across different cultures. More convincingly, he has used objective measurements of facial behaviour (how much different parts of the face move) and compared these across cultures and countries – for instance, by testing participants from Japan and the USA (Ekman, 1973). His studies have also extended to infants from different cultures (Ekman and Oster, 1979). You may wonder why Ekman chose to look particularly at infants' facial expressions rather than adults. If a characteristic or ability is present in infants, who have had little opportunity to be influenced by their culture or upbringing, then that is additional evidence for that characteristic being largely genetic rather than learned. The spontaneous facial expressions of blind children (Medicus *et al.*, 1994; Eibl Eibesfeldt, 1988) also supports the idea that there may be basic emotions, and further that they may have biological rather than social origins. Other theorists such as Plutchik and Frijda (Plutchik and Landau, 1973; Frijda, 2001; Frijda and Tcherkassof, 1997) rely not only on facial expressions but on whole body movements – what is often called 'body language'.

Ekman's assumption about the inheritance of emotion is shared by many others promoting the notion of basic emotions. For these researchers it follows that such emotions arise from subcortical brain mechanisms that we still share with many other species (e.g. Panksepp, 1989; Panksepp *et al.*, 1991; and LeDoux, 1989). Debate is still active concerning whether and which emotions are basic, or whether it is clusters of related emotions that should be considered together. However, many believe that the development of the brain systems underlying something approximating to the basic emotions seems to have arisen far back in our evolutionary past before the separation of mammals, reptiles and birds. Box 13.2 discusses how technology for imaging the brain has begun to shed light on the relationship between specific emotions and particular brain structures.

## 13.2

## Methods

### Imaging emotions in the brain

Although techniques for imaging the structure of the brain (taking pictures without the need to make any actual physical intrusion) have been used in medicine for several decades, the ability to study changes in activation associated with brain function is a more recent and rapidly growing technique. Two common techniques are PET (Positron Emission Tomography) and fMRI (functional Magnetic Resonance Imaging). PET involves injecting the participant with very slightly radioactive water, which then travels around the body including the brain, emitting its radiation as it goes. Participants perform an experiment usually designed to test particular hypotheses about which brain areas are involved in the task(s) concerned. Because the most active areas of the brain will draw the most blood, these areas will also be emitting the most radioactivity. This is measured using special gamma ray (the energy component of radioactivity) detecting equipment, and thanks to complex software can be translated into brain images, which look similar to those shown in colour Plates 6 and 7. fMRI also produces



similar looking images of brain activation involved in a task – but using a different method. A strong magnetic field is applied across the brain that aligns certain particles in the blood (called ‘de-oxygenated haemoglobin’ molecules) in the same direction (similar to the way iron filings line up towards a magnet). When this field is removed these particles ‘precess’, or move back again, and in doing so each particle emits a discrete ‘package’ of energy, which is detected by specialist equipment. The more active an area of the brain is, the more of these particles it has and consequently the more energy is emitted during precession.

Brain-imaging techniques such as those described above are revealing some very interesting results about emotions. For example, many studies have now shown that a structure called the amygdala is involved in the processing of all types of emotion and is particularly strongly activated in response to fear stimuli. Similarly, two areas are implicated in recognizing disgust: the insula, an area of cortex (the convoluted outer layer of the brain) and the basal ganglia (an evolutionarily old area in the brain stem). This has been corroborated by data from a patient who has damage to these areas and is particularly poor at recognizing disgust in others (Calder et al., 2000).

Colour Plate 6 shows the areas of the brain where different studies have reported activation resulting from either the processing of fearful faces (green squares) or learning about fear (red circles). The image on the left is a horizontal ‘slice’ through the brain, with the eyes at the top end and the back of the head at the bottom. The image on the right is the same sort of slice, but taken higher up, more towards the top of the head. There is a tendency for the activation triggered by processing fearful faces to involve the left amygdala, whereas learning about fear seems to produce more bilateral activation.

Colour Plate 7 shows the areas where studies have found brain responses to disgust. As with Plate 6, the two images depict different slices through the brain. The insula activations are shown in purple, and basal ganglia activations in red. The basal ganglia signals are mainly in the right hemisphere, whereas the insula signals are more evenly distributed across the two hemispheres.

Another feature of these new brain-imaging techniques is that as well as the cognitive processing of emotion described above, they can give us an objective measure of the ‘feelings’ side of emotion (see Section 1.1.3). As we said earlier, feelings have been notoriously hard to study in psychology because the only way to measure them was to rely on people’s subjective self-reports of their own internal state – the much-scorned ‘introspection’. Now though, we can investigate how brain activity changes according to the strength and nature of our feelings and this is a possibility that is only just starting to be exploited.

## 2.2 Verbal labels

The fact that very similar verbal labels are used across widely differing languages and cultures is sometimes used as evidence in support of the existence of a discrete set of basic emotions corresponding to those labels. Scherer and colleagues (e.g. Scherer and Wallbott, 1994a and b; Wallbott and Scherer, 1988, first published

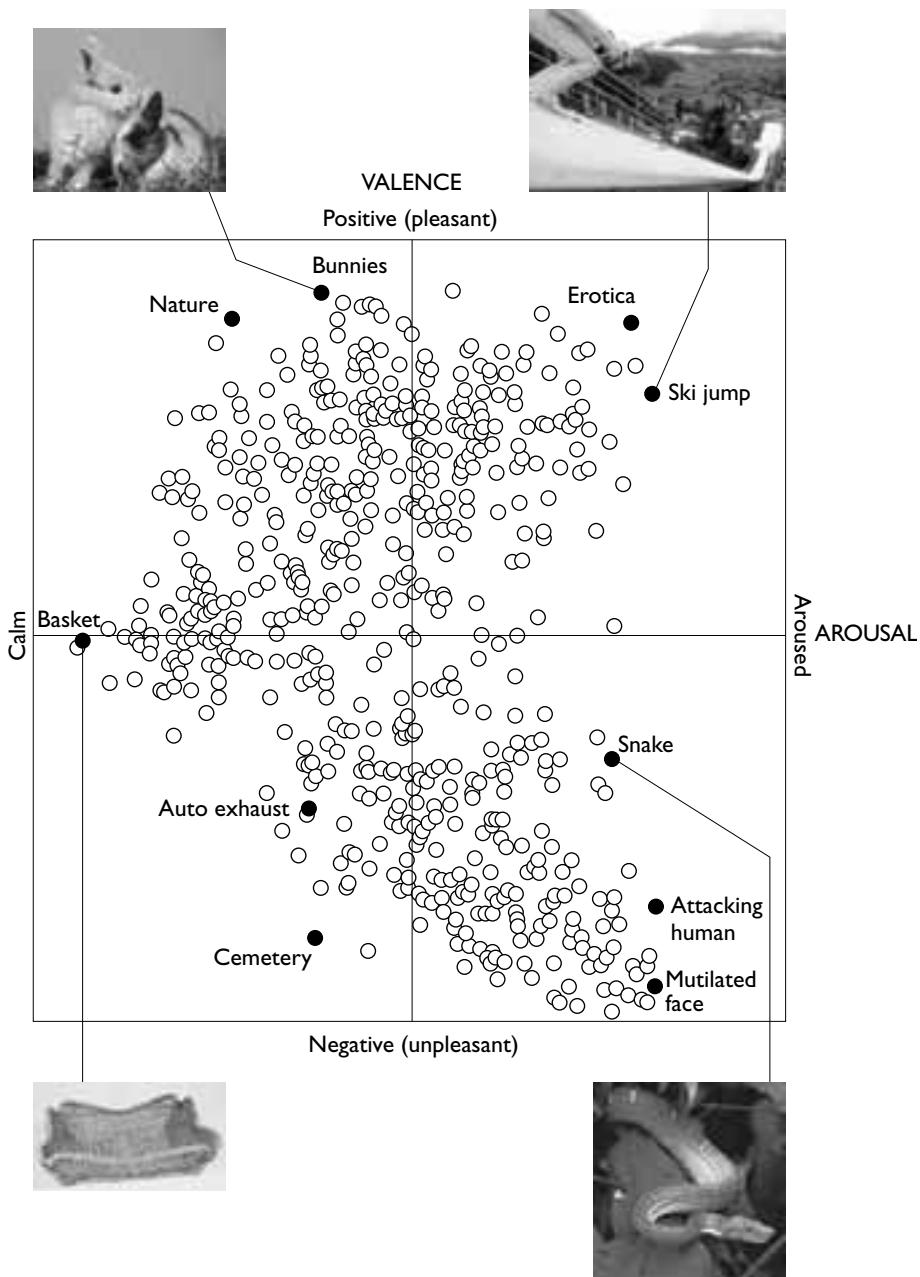
1986) have compared verbal labels for emotions in 37 countries and were able to translate the English terms for the seven emotions studied (anger, fear, sadness, joy, disgust, shame and guilt) into each of the other languages. If all languages include words to describe the so-called basic emotions, and these emotions can be recognized across all cultures, however remote or different from each other, then that gives reason for believing in the universality of the concepts for the basic emotions. What about all the other emotion words, where do they fit into the idea of basic emotions? Scherer and others introduce the idea of ‘modal emotions’; that is, the idea that a number of these other emotion words may cluster together under a common ‘theme’, and that the specific clustering of emotion words betrays the underlying emotion concepts of the individual (recall discussions about concepts in Chapter 5). To complicate matters, different languages and cultures do seem to differ in the number and categorization of their emotion terms. It is not surprising to find that the range of situations that trigger emotions varies across cultures, but, in addition, different emotions are either elaborated or downgraded in emphasis. There appears to be a set of universals – for instance, loss of a loved one leading to sadness, and attack to fear or anger – as well as a multitude of cultural specifics, such as whether looking directly at a woman’s face evokes sensations of polite interaction, flattery or insult. Whilst debate continues about whether there are a small number of basic emotions and whether these are necessarily inherited, it is clear that there are many cultural differences in emotions. Thus, there are cross-cultural differences in:

- the number and type of complex emotions
- the triggers for many emotions
- the socially acceptable rules for which emotions should be displayed in certain contexts.

### 2.3 The dimensional approach

The concept of ‘basic emotions’ is not without challenge. Theorists such as Ortony and Turner (1990) have asked why, if basic emotions are so basic, there is so much disagreement about which count as basic, with some contenders (e.g. interest and desire) sometimes not even being considered as emotions at all. An alternative dimensional approach, as the name implies, assumes that the full range of emotional experience can be explained by identifying a few key dimensions. If there are only two key dimensions, then all emotions could be identified as being located in a two-dimensional space specifying the relative contribution provided by each of the two dimensions. An example of this approach is shown in Figure 13.4.

The ‘affect grid’ in Figure 13.4 is taken from work by Peter Lang and colleagues, who concentrate on studying our physiological responses to emotional material. There are two dimensions, **arousal** and **valence** (valence refers to the positive/pleasant or negative/unpleasant qualities of something). The figure shows people’s ratings of how ‘aroused’ and how ‘positive or negative’ they feel about a variety of different pictures. Other, separate dimensions, such as ‘dominance’ have also been proposed, producing a more complex three-dimensional space (dominance reflects a quality related to how dominated vs in control the participant feels when considering



**Figure 13.4** The affect grid and example pictures (the ratings on the grid are for pictures similar to those shown)

Source: Dawson et al., 1999, Figure 8.2, p.161

this emotion or emotional information). This approach circumvents some of the problems associated with basic emotions. It has the advantage of suggesting how the different emotions relate to one another and makes it easy to understand how different languages could have developed different words to describe subtly different mixes of emotion experience. However, some emotions appear to combine

attributes that the dimensional model assumes should be at opposite ends of a single continuum. For instance, nostalgia seems to combine attributes of positive valence (the positive value of past experiences) with negative valence (sadness or regret at their passing); and the excitement of extreme sports or roller-coaster rides combines pleasure (positive valence) with fear (negative valence) to create the characteristic exhilaration and excitement. Furthermore, it is still necessary to determine the dimensions, and how many should be used, and to decide how these relate (if at all) to the evidence suggesting the existence of basic emotions.

### ACTIVITY 13.3

---

You might like to consider how you would map the different discrete emotions onto the affect grid. Where would you place sadness, contentment, fear and excitement, for example?

#### COMMENT

You have probably opted for bottom left, top left, bottom right and top right for sadness, contentment, fear and excitement respectively. Sadness, for example, could be considered as fairly unpleasant with little excitement or energy. But notice too that the grid allows for a lot more variation between items. You may also notice that the distribution has a 'C' shape to it. It appears that there are plenty of things that we consider to be neutral (neither pleasant nor unpleasant) and not particularly arousing, but very few arousing neutral items! Putting it another way, if something is arousing we tend to find it either really good or really bad (or as already noted, maybe both together).

---

## Summary of Section 2

---

- Some psychologists classify different emotions by identifying discrete or basic emotions.
- The Big Five basic emotions are widely recognized by many psychologists and there is reasonable evidence to support this classification.
- The use of similar verbal labels, and production and recognition of emotional expressions across different cultures further supports the notion of basic emotions.
- An alternative approach is to use a small number of continuously varying dimensions to describe the range of emotional experience.

## 3 The function of emotions

Emotions and emotional responses to events could surely not have evolved unless they served a useful purpose, but what purpose or purposes might these be?

**ACTIVITY 13.4**

Can you think of any aspect of emotions already touched upon in this chapter that might bestow a useful advantage on animals (including humans)?

**COMMENT**

Think (or look) back to the Section 1.1.2 on physiological responses to emotion. You will recall that, in response to a frightening event, rapid physiological changes take place that prepare the body for 'fight or flight'. Undoubtedly the rapid mobilization of the body's resources in this way provides a potentially life-saving advantage.

### **3.1 Emotions alter goals**

One influential modern theory of the function of emotions is that of Oatley and Johnson-Laird (1987). They have proposed an evolutionary account of emotions that suggests the role of emotion is to signal that ongoing behaviour should be interrupted to take account of a conflicting goal. They argue that humans have many different motivations and goals. Events will happen that require setting or resetting of priorities amongst these goals, such as giving up the goal of planting next summer's food crop in favour of running away from an attacking lion. For example, sadness caused by bereavement is not maladaptive, but in their framework is seen as having the function of initiating readjustment of life goals that included the lost one. When the relationship was close, this period of reassessing or reforming goals could be lengthy.

**Table 13.2** Summary of emotions and their associated goals according to Oatley and Johnson-Laird

<b>Emotion</b>	<b>Juncture of current plan</b>	<b>Behaviour/response</b>
Happiness	Subgoals being achieved	Continue with plan, modifying as necessary
Sadness	Failure of major plan or loss of active goal	Do nothing/search for new plan
Anxiety	Self-preservation goal threatened	Stop, attend vigilantly to environment and/or escape
Anger	Active plan frustrated	Try harder, and/or aggress
Disgust	Gustatory goal frustrated	Reject substance and/or withdraw

Source: Oatley and Jenkins, 1996, Table 9.1, p.256

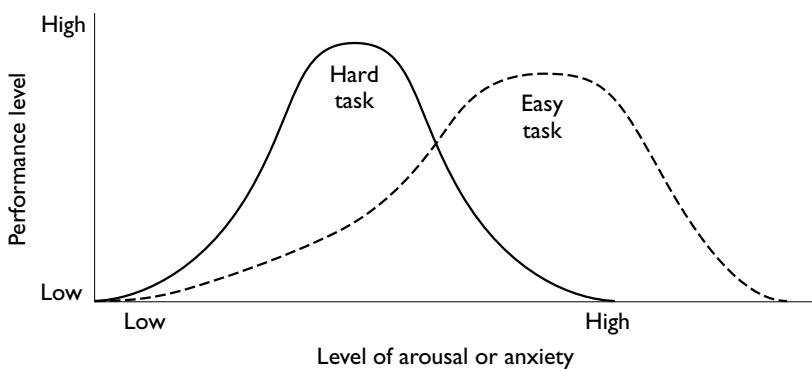
Central to Oatley and Johnson-Laird's theory is the notion of cognitive readjustment to emotional events. However, the exact mechanisms are not spelled out. Unlike the physiological changes that we discuss next, the mechanism behind changes in cognitive processing in response to emotions is much less well understood. Later in the chapter (Sections 4.2 and 4.3) we touch on some aspects of attentional deployment and memory in emotion, but there is still a considerable shortfall in our understanding of how internal emotional status influences cognition and how processing of emotional information is prioritized and influences cognitive function.

### 3.2 Emotions mobilize physiological resources

It is relatively easy to see how the physiological changes involved in emotions are part and parcel of the need to readjust goals, sometimes with great rapidity. It is vital that if your life is threatened, then your body is ready to respond in the best possible way to ensure your survival. In Section 1.1.2 we described some of the bodily responses associated with emotions such as fear and gave an example of how these assist in ensuring survival. The physiological reactions described there stand us in good stead when physical action is required. Often in modern life, however, an emotional threat requires not increased physical exertion but less. One example is the threat of a pending examination, which requires long hours sitting still at a desk to revise rather than any physical exertion of the body. Similarly, most of us have experienced the fear of a near accident while driving, but all our bodies actually need to do to avoid the danger is perform minimal, albeit rapid, movements of the hands on the steering wheel and the feet on the brakes.

Does this mean that many emotional reactions, especially fear responses, no longer have useful functions? No. There are still many occasions when rapid physical responses avert death or injury. Even before an exam, when you won't be fighting or fleeing (even if you feel you'd like to), the increased adrenalin and physiological arousal will provide an energizing effect that can improve performance, if maintained at an optimal level. However, too much anxiety impairs performance as the anxiety itself interferes with cognitive function and the physiological reaction makes it hard to relax and sit still. At the other end of the spectrum, not enough arousal – in other words boredom or disinterest – also impairs performance.

This finding, expressed formally by Yerkes and Dodson (1908), is known as the **Yerkes–Dodson law**, and is shown in Figure 13.5. Notice too that, for an easier task, higher levels of arousal are needed to attain the optimal level of performance compared with a hard task. The Yerkes–Dodson law seems reasonable and suggests that appropriate levels of emotion can indeed be useful – both too much and too little emotional arousal can put an organism at a disadvantage.



**Figure 13.5** Yerkes–Dodson law

Source: Eysenck and Keane, 1996, Figure 18.8, p.454

### 3.3 Emotional expressions as communication

What function might emotional expressions fulfil and can we see evidence of them in other animals? Charles Darwin formalized the evolutionary view of the emotions in one of his later works entitled *The Expression of the Emotions in Man and Animals* (1998, first published in 1872). As well as acknowledging the more obvious evolutionary advantage of physiological changes during emotions (such as preparation for fight or flight), he highlighted how expressions of emotion serve to communicate the emotional status of an animal to others of their species (so-called conspecifics). However, it has to be said that he felt that emotions and emotion expressions in humans were no longer functional, but merely a relic from our evolutionary past – much as our appendix is seen to be superfluous to digestion. He drew parallels between the expressions of animals and their functions, such as the snarl of a dog communicating a readiness to bite, and the sneer of a human, which presumably has the same origin but no longer sends the same message (normally!).



**Figure 13.6** Darwin's comparison between the sneer of a woman and the snarl of a dog  
Source: Darwin, 1998, Figures 14 and 22, pp.117 and 246

Darwin might have been partially correct in his feeling that emotion expressions no longer have the function in humans that they do for other animals. We have already considered that display rules for expressions differ from culture to culture. For this difference to occur, then of course we need to control the expression displayed (at least to some extent). This means also that humans are capable of deceiving with their expressions – we can lie about our emotional feelings. Thus, emotional expressions serve multiple functions for humans: they can be honest signals of emotional status, as in other animals, or they can be part of the impression management, polite interaction or social manipulation of the sender.

### 3.4 Emotions as information

To illustrate the idea of emotions as information let us return to the Capgras delusion first mentioned in Chapter 4 (Section 6, Box 4.2). To remind you of the delusion (or syndrome) we will illustrate it with the case of an individual whom we will

call Alan. Alan was in a car accident with his wife. He sustained an injury to his head, and his wife, Christine, was also injured, taken to hospital but later recovered. Alan refused to believe that Christine was still alive. He recognized her face but remained convinced that this was not really her, but a sinister impostor. Remember that in this rare syndrome the sufferer believes that a family member, or someone close, has been replaced by aliens or impostors. In such cases it is believed that although facial recognition is intact, a parallel system for registering the emotional meaning of the face has been damaged. Indeed, when Alan's palms were tested for SC changes (skin conductance – see Box 13.1) when viewing pictures of Christine's face, these responses were absent. SC changes, signalling an emotional response, would normally occur for any of us if viewing either emotional expressions or the face of someone we know. Without them there is no emotional resonance, no sense of affiliation. It would appear that Alan's brain interpreted this lack of physiological feedback as evidence that this was not someone close. However, since the perceptual qualities of the face matched those of his wife, it must be someone who looked just like Christine, an impostor, a frightening and distressing situation for all.

The idea that emotions provide information to guide decision making is fundamental to the theories of Damasio (1996). His views are best explained by describing the task most associated with him, the so-called **gambling task**. You are given four decks of playing cards and asked to select from one and turn the card over. For reasons that are not explained, you are either rewarded or fined as a result of your selection. Your task is to attempt to maximize your winnings. After playing for a while you are likely to find yourself making more selections from two out of the four decks, but you probably won't be able to say exactly why. This task is arranged so that two decks (the 'good decks') give less spectacular wins but also less punishing losses. Choosing from these two over a period of time achieves a modest gain. The other two 'bad decks' sometimes deliver large wins, but also large losses resulting in an overall loss on average. The rule is not hard and fast and so is not generally very obvious as you play. Playing the 'good decks' is the best strategy; you win a little and lose a little, but overall your winnings start to add up. Damasio has developed a theory to explain how people come to operate this strategy successfully, and make other similar decisions in life, when they are acting on hunches rather than full understanding.

According to Damasio, the emotional responses to winning and losing produce physiological changes that he calls **somatic markers**. Over time, through a process of conditioning, the decks come to evoke different physiological responses in the player, essentially representing the accumulated positive emotions of wins together with negative emotions of losses. After extended experience, as the player considers making a selection from each deck the physiological response conditioned to this deck will be initiated and this acts as a marker capable of guiding choice. Damasio suggests that somatic markers represent the 'gut feelings' that we often use to guide our decisions even though we may never become consciously aware of why we have a gut feeling about a particular choice. The function of emotion, for Damasio, is therefore centred around information and future actions.

### 3.5 What is the function of emotional feelings?

Although we have only touched on the topic, it is relatively easy to propose evolutionary advantages conferred by emotional behaviours and physiological responses to emotions. The same is not true of the function of emotional feelings. Take fear, the example we have used the most. The physiological response to a fearful situation can provide the physical resources to escape danger or stand and fight. Our behaviour, including expressions, functions to communicate our fear to others. But why do we need to experience the unpleasant *feeling* of fear, or anticipated fear (anxiety) when we expect a frightening experience?

Feelings are part of our conscious experience. The functions of consciousness, as you will consider in Chapter 15, are by no means uncontroversial, but one facet that is fairly regularly acknowledged is the notion that consciousness is necessary for performing new tasks or trying to override habits without relying simply on mechanisms of conditioning. This would also apply to learning new responses to an emotional situation such as when soldiers continue to advance into battle despite a strong urge to flee, or in overriding habits such as when suppressing the tendency to respond in anger at a socially inappropriate moment. However, whilst these examples invoke the need for consciousness, they still do not explain the necessity to *feel* the emotions of fear or anger. As the psychology of emotion continues to develop, future theories and research are likely to give us greater insight into the possible function of the feelings associated with emotion.

## Summary of Section 3

---

- Following Darwin, many psychologists believe that emotions have evolutionary functions, including the mobilization of physiological resources, which remain today.
- Oatley and Johnson-Laird maintain that the purpose of emotions is to interrupt current behaviour in order to change priorities and goals in the light of new information.
- Damasio has formulated a somatic marker account of the function of emotions, in which their primary purpose is to provide information, via bodily feedback, which guides future decision making.
- The functions of ‘feeling’ emotion are still speculative.

## 4 Emotion influences cognition

### 4.1 Some important concepts

Before we start to discuss how cognition and emotion interact with each other, there are some important distinctions that you need to become familiar with. The first of these is the distinction between trait and state emotion.

#### 4.1.1 State and trait emotion

**State emotion** (also called mood or affect) refers to how you feel right now. As you will be aware, this can change from minute to minute, day to day. State emotion is a very transient and variable entity. It is a construct that allows us to acknowledge the fact that momentary feelings may be quite different from the way an individual usually feels. Although state emotions are usually measured by self-report (asking participants to introspect and describe how they feel), they also relate directly to the behaviours and physiology discussed above and can be measured in the same way.

In contrast, **trait emotion** refers to more stable personality characteristics or ‘what kind of person’ you are. For example, some individuals may be prone to angry outbursts, or have a tendency to worry about things, or be optimistic, always looking on the bright side. Psychologists have directed much effort into trying to capture theoretically these ideas about stable personality characteristics. Thus, traits are theoretical constructs relating to aspects that are more enduring and characteristic of a person, and describe how one person may differ from others. Some common traits that have been proposed and are frequently measured (again by self-report) include: anxiety; depression; social desirability (how much you adapt your behaviour in order to gain the approval of others); anger; impulsivity; and emotional sensitivity.

A trait tends to make a person more prone to experiencing the associated mood state. For example, a high trait anxious individual will tend to feel more anxious for more of the time than a low trait anxious person. This is why certain traits, like anxiety or depression (sadness), are useful to psychologists interested in emotions – they are a more permanent indicator of who tends to have more or less of the relevant state emotion.

#### 4.1.2 Processing vs manifestation of emotion

Another important distinction is between the *manifestation of emotion* itself and the *processing of emotional material*. The manifestation of emotion is exactly what we were discussing in Section 1. Thus, by ‘manifestation’ we mean both the experience of emotion, the feeling state, and the expression of that experience through bodily changes and behaviours. This is also often known as the ‘hot’ component to emotion. In contrast the ‘cold’ component is the processing of emotional material but without emotion being actually experienced. This isn’t always an easy distinction to make. It is a bit like the difference between describing an emotional event in a detached way (relating a series of facts) compared with describing it in emotional terms. Obviously the two types of process regularly co-occur – the memory of the facts of an event often brings back the feelings as well – and in this case the manifestation/processing distinction may seem blurred. However, in psychology we often use stimuli such as words or pictures as a way of studying how we process emotional material although these stimuli rarely elicit a strong experience of emotion in participants. It is important to grasp then that studying cognitive processing in emotion can be quite distinct from studying the manifestation of emotion.

You may already have realized that the processing of emotional material is our first example of an interaction between cognition and emotion. In a typical experiment one might present participants with lists of negative emotional words (e.g. cancer, attack, evil) mixed with neutral words (e.g. number, unusual, round) and ask for later recall in a surprise memory test. The emotional aspect in the task is the valence (pleasantness/unpleasantness) of the words, which is the independent variable. The cognitive measure (dependent variable) is how many words of each type are recalled in the memory recall test. You might be interested to know that while most people will remember more positive than negative words (a very common ‘positive bias’ in emotion processing), individuals with clinical depression tend to remember more of the negative words. Box 13.3 in Section 4.2.1 discusses this further.

As we mentioned above, things can become more complicated when hot and cold emotions occur simultaneously. In the psychology of cognition and emotion we are interested not just in how people process emotional material, but also in how this processing is affected by emotional states and traits. For example, does the processing of sad words change when someone is actually feeling sad at the time? Similarly we might want to know whether people who are vulnerable to anxiety (i.e. high on trait anxiety) process threatening words any differently from those who are not. These more complex questions are what cognition and emotion psychologists are mostly concerned with. Section 4.2 delves deeper into these issues.

## 4.2 Memory

We start our examination of the interactions between cognition and emotion by considering the ways in which emotional states affect memory for emotional material.

### 4.2.1 Mood congruent memory

What happens to memory processes when the content of material being encoded matches the mood state of the participant doing that encoding? For example, if you are feeling sad and then happen to watch a sad film, how does this influence your later memory for the film? This scenario could produce an example of **mood congruent memory (MCM)**. Bower and colleagues’ classic experiments sparked a great deal of interest in this phenomenon. In a typical example, participants are put in either a happy or sad mood by hypnosis and then read both a happy and a sad story (you may like to consider the ethical implications of doing such a study). Participants are then given a surprise recall test to see how much of each story they recalled. The results are shown in Figure 13.7 overleaf.

As you can see, more was recalled from the story which matched the mood of the participant as they were reading; for example, sad participants recalled more things about the sad story than about the happy one. The phenomenon of mood congruent memory has proved very robust. Also it has sparked a whole field of research into the effects of emotional disorders on cognitive processing, such as the relationship between clinical depression and memory processes.



**Figure 13.7** Results from Bower's (1981) mood congruent memory experiment

Source: Eysenck and Keane, 1996, Figure 18.4, p.446

### ACTIVITY 13.5

You might like to think about what you would predict if you tested a clinically depressed person on memory for negative and neutral information, bearing in mind that one of the hallmarks of clinical depression is chronic low, or sad, mood. After you have considered this, look at Box 13.3.

Interestingly, work on mood congruency has alerted us to the finding that even 'normal' individuals, in no particular mood, seem to have a positive and potentially adaptive bias towards memory for positive information. Some suggest that this helps us to keep a positive outlook on life, in the face of all the problems it throws at us. It is as if we are 'looking at the world through rose-coloured glasses'!

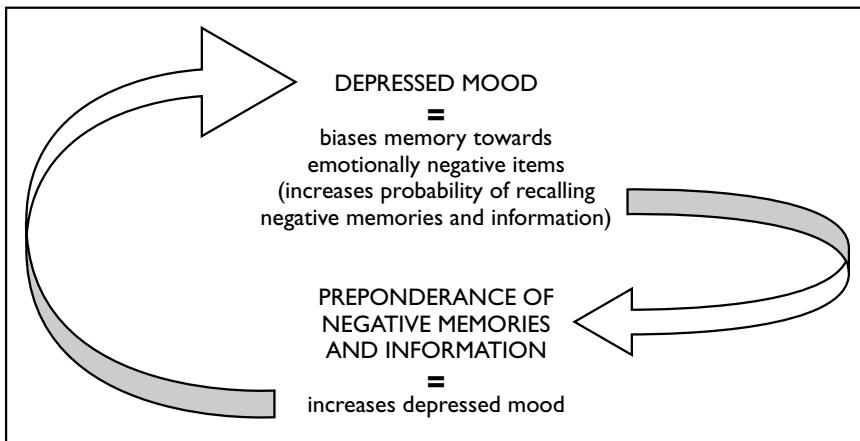
#### 4.2.2 Mood dependent memory

**Mood dependent memory (MDM)**, or mood state dependent recall is a well-known, but controversial phenomenon. It can be seen as a specific case of the influence of context on memory that was described in Chapter 8. The idea is that your memory for a particular stimulus or event will be better if there is a match between your mood at the time you experienced it and your mood when you try to recall it. For example, imagine you have a heated argument with a friend. Mood dependent memory would suggest that you will remember more of what was actually said if you are in an angry state again than if you are not.

### 13.3

#### Clinical depression and memory bias

Typically, individuals with clinical depression and those who are not diagnosed but still report feeling constantly low in mood (sub-clinical depression) all show mood congruent memory (MCM) effects, sometimes called a 'bias', for negative material. Many different types of experiment have been used to verify this finding, using positive and negative word lists, self-descriptive adjectives, sentences and whole scripts (Matt *et al.*, 1992 offer a meta-analysis). The effect appears to be stronger when participants are aware of the relationship between their mood and the material; and, not surprisingly, when the negative nature of the material is stronger (e.g. 'evil' vs 'bad'). The bias also includes recall of autobiographical memories (see Chapter 14). Although this method might seem inconclusive (maybe depressed people really have had more negative experiences anyway), experiments using mood induction really do suggest that mood affects the valence of the personal memories that are brought to mind.



**Figure 13.8** It is thought that the mood congruent memory effect contributes to a vicious cycle in which depressed mood enhances the accessibility of negative memories. In turn, having more negative memories in mind is likely to exacerbate depressed mood

Source: based on Teasdale, 1988

These findings are of more than just theoretical interest. It has been suggested that MCM may contribute to keeping someone in a depressed mood and that if we change this cognitive processing bias, then that might help the mood to lift. Teasdale (1988) has developed this idea, as shown in Figure 13.8. The suggestion is that patients' bias towards recalling more negative mood congruent information means that their world will seem more full of negative things than is really the case. This in turn will make them feel even more depressed. You can see that a vicious circle could be set up, where the memory bias contributes to the mood, which enhances the memory bias and so on. Teasdale and others have spent a lifetime of research trying to devise methods of breaking this cycle and coming up with new cognitive treatments for depression, such as a procedure called mindfulness-based cognitive therapy (Segal *et al.*, 2002).

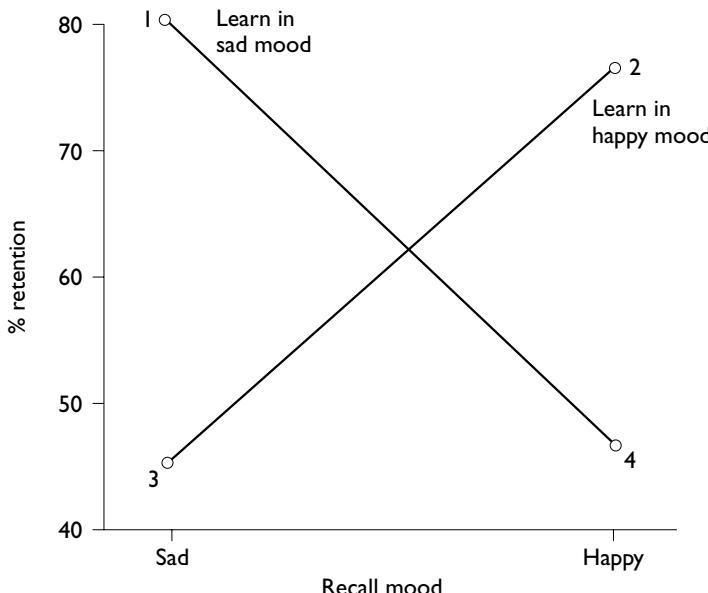
In the laboratory this hypothesis has been tested using the following type of experiment. Participants are put into particular moods (mood induction) by one of several techniques such as hypnosis, listening to appropriate music or reading appropriate passages of text. Then they are asked to learn a list of arbitrary, neutral words while in the induced mood. Participants are later put back into either the same or a different mood and asked to free recall the words ('remember as many as you can'; no cues or prompts are given). If this second induced mood matches the one they were in when they learned the first list, then recall should be higher.

A classic experiment of this type is that by Bower (1981) who used happy and sad mood induction by hypnosis. Figure 13.9 shows some typical results from their experiment. In this design participants learned two lists of words, list A and list B, one after the other, but only recall for the first list, list A, was tested. As usual, participants were put into either a happy or sad mood before learning took place, one mood for each list. Thus those who learned list A in happy mood then learned list B in sad mood, and vice versa. Then, after both lists had been learned, they were tested on their recall for just the first list. The mood of participants during the test (using a third mood induction) either matched or contrasted with the mood at the time of learning list A. So for some participants mood at recall matched mood at learning (points 1 and 2 in Figure 13.9), whereas for others mood at recall was different from mood at learning (points 3 and 4 in Figure 13.9). You should be able to see from the figure that when learning and test moods were the same, participants were indeed better at remembering list A, compared with participants who tried to recall the same list in a different mood from the one they had learned it.

Perhaps you are wondering what was the point of the second list B? The reason for using two lists was simply that learning list B in a contrasting mood acted as an interference task, which made the experiment more sensitive to the beneficial or detrimental effects of the mood manipulations.

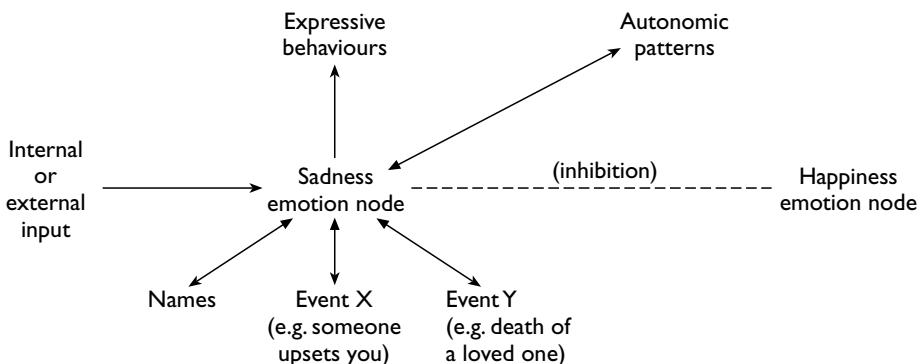
Bower (1981) went on to propose an influential **semantic network** theory to explain these mood and memory effects. The theory is shown in Figure 13.10. Bower suggested that emotions could be represented as nodes in a network, having numerous connections to related semantic items (words, concepts, etc.), other emotion nodes and outputs such as behaviour and autonomic responses. Material such as memories and knowledge is stored in the network and may be connected to some emotion nodes. Nodes become activated by external or internal stimuli and when this happens that activation selectively spreads across the network via the links to other units, a bit like ripples across a pond. Notice that some connections are inhibitory, so that activation of the sadness node, for example, would suppress any activation in the opposite happiness node. When nodes are activated above a certain threshold, then the content of those nodes enters conscious awareness leading to the corresponding feelings and thoughts.

You can perhaps start to see how this theory fits with the results of mood dependent memory experiments. When participants learn a word list in one mood, links are created between the relevant emotion node and the memory representations of those words. Thus when participants try to recall the same words this can be made easier if they are in the same mood thanks to the spreading activation from the



**Figure 13.9** Percentage retention of words according to the match between learning mood (happy or sad) and recall mood

Source: based on Bower, 1981, Figure 2, p.132



**Figure 13.10** An example semantic network theory of emotion

Source: Power and Dalgleish, 1997, Figure 3.2, p.71

associated emotion node. Conversely, in a different mood there will be no advantage from such activation, and indeed it is assumed that inhibition of the word representations from the incongruent emotion node might result.

Since Bower's original experiment there have been many attempts to replicate his finding, but these have met with varied success. It seems that mood dependent memory is not a very robust effect. It is much influenced by factors such as the strength of the mood state that is induced, and the nature of the items to be recalled (e.g. recall using real-life autobiographical events produces better results). However,

in a recent review of the work on mood dependent memory Eich and Metcalfe (1989) concluded that the phenomenon itself was genuine, and that the problems lay with the methods used to detect and measure it. There is no doubt that Bower's findings and theory have been remarkably fruitful in their influence on the thinking and direction of emotion research.

Before we move on, it is worth stopping to think what the key difference is between mood dependent memory experiments and those we discussed in Section 4.2.1 under the heading mood congruent memory. Here we have been concerned merely with the effect of mood on recall, irrespective of what it was that was actually being remembered. With mood congruent effects however there is always a match – or congruity – between the emotional material being recalled and the mood of the individual when encoding that material. Congruity means a match between mood at encoding and material being encoded; dependency refers to a match between mood at encoding and mood at retrieval.

### ACTIVITY 13.6

---

Can you think of examples where congruent or incongruent stimuli (rather than mood states) might influence cognition?

#### COMMENT

It has regularly been shown that an individual's performance is influenced by whether or not two separate aspects of a situation are matched or not. For example, in tasks demonstrating the Flanker effect (see Chapter 2, Section 3.3) you may remember that performance depends on how closely matched the targets and the distractors are. The Stroop effect, discussed in Section 4.3 below, is another such example.

---

### 4.3 Attention

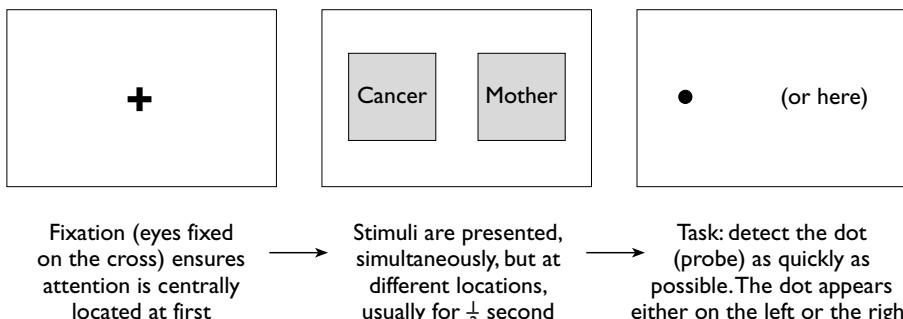
In the same way that memory for emotional material can be biased in a direction consistent with one's mood, so can attention. A classic example of this is the 'emotional Stroop'.

In the standard Stroop task (Stroop, 1935) (see Chapter 2, Section 3.3, Box 2.2), participants are asked to name out loud, as fast as they can, the colour of the ink in which colour words are written. When the ink colour is different from the meaning of the word itself (e.g. 'blue' written in red ink) participants are slowed down compared with stimuli where the word meaning and ink colour are matched. The effect arises because of the different amounts of interference between congruent ink colour and word meaning compared with a competing or incongruent ink colour and word meaning. The emotional Stroop differs in that, instead of colour words, emotional and neutral words are used, still printed in different colours. Examples of both types of Stroop task are shown in colour Plate 8.

When the emotional Stroop is given, for example, to high trait anxious individuals, then the interference from the anxiety relevant words is usually greater than that from the neutral words, compared with the same difference when observed in non-anxious individuals. As performance on the Stroop task is generally taken to

be a measure of attention towards the word meanings (although the precise mechanisms behind the effect are still not fully understood), then this is an example of an *anxiety-related attentional bias*.

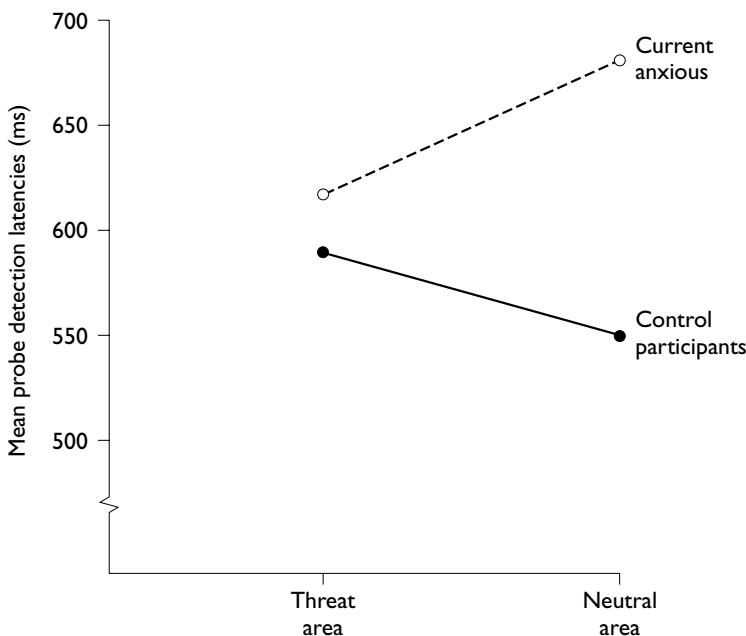
In an attempt to demonstrate more clearly the nature of this attentional bias, MacLeod *et al.* (1986) published a now classic paper using an innovative new method of testing attention allocation. Their design, now known as the **dot probe** or ‘attentional probe’ task, is shown in Figure 13.11.



**Figure 13.11** The dot probe task

The task is to respond as rapidly as possible to the presentation of a dot (termed a ‘probe’ because it is probing where attention is located). This is, therefore, a reaction time (RT) task. On some trials (catch trials) there is no dot, to make sure participants are really looking for it and not just responding as soon as the words disappear. As you can see, before the dot a pair of word stimuli are displayed, one threatening and one neutral. If a participant is consistently faster to find the dot whenever it appears where the threatening item was, then we can reasonably assume that they must have been attending to that item rather than to the neutral item. The original results of MacLeod *et al.* (1986) are shown in Figure 13.12 overleaf.

The figure shows that control (not anxious) participants were just slightly faster when probes appeared in the neutral rather than the threat areas of the display (another example of the normal ‘positive bias’). Anxious patients were the other way round – faster for probes appearing where threat words had been than for probes appearing where neutral words had been. This strongly suggested that anxious individuals allocate their attention to threat words rather than to neutral words, whereas controls do not. Thus, consistent with the emotional Stroop results, MacLeod *et al.* found an attentional bias for threat in their anxious patients. These results sparked over a decade of continuing research into this so-called attentional bias for threat. We now know that the bias is seen with many different types of material including words, pictures and faces, but is most prominent when the material matches the current concerns of the individual. For example, snake phobics will show a stronger attentional bias towards pictures of snakes than towards pictures of snarling dogs. This type of bias, with suitable materials, has been shown with patients suffering a variety of anxiety disorders, such as those with phobias, generalized anxiety, and post-traumatic stress disorder. It is also apparent, although



**Figure 13.12** Results from MacLeod *et al.* (1986)

Source: Eysenck and Keane, 1996, Figure 18.6, p.450

less reliably so, in non-clinical individuals who have high state anxiety or high trait anxiety, or both.

Look again at Figure 13.8 in Box 13.3, where it is suggested that depression could be exacerbated by a vicious cycle of mood congruent memories contributing to sad mood. Could a similar mechanism be involved in attentional processing? Mathews (1990) proposed that just such a vicious cycle could operate to maintain anxious mood and attention to threat. Imagine that your anxiety makes you pick out and pay more attention to potential threats in the environment. This bias may well make it seem as if your surroundings are full of threats and this would, unsurprisingly, make you feel more anxious, which would perpetuate your attentional bias and so on. You would end up in a hyper-vigilant state, anxious about, and on the look out for, threats of relevance to you. This idea that anxious people are constantly in a vigilant, checking mode is popular in current theorizing about clinical anxiety.

How could you test out whether an attentional bias might cause anxiety or be caused by anxiety (or both), and can anything be done to reverse its effects? Mathews and MacLeod (2002) set themselves such a task by devising methods of directly inducing a positive or negative bias in non-anxious volunteers using specialized training procedures, and then assessing its effects on anxiety levels. Possessing an induced positive bias reduced, whilst a negative bias increased, anxiety levels when exposed to a moderately stressful situation just after training. These results certainly confirm that the attentional bias has a causal effect on anxiety levels, and interestingly that training procedures have been found that directly modify the bias, and can thereby reduce (or increase) anxiety.

Another similarity between attentional bias and the memory biases we discussed in Section 4.2 above is the performance of ‘normal’, non-anxious controls. As before, it seems that most of us have an adaptive or protective bias in the opposite direction to that of emotionally disordered patients. Look at Figure 13.12 again. Controls are faster in neutral areas than in threat areas, and this has also been found in several subsequent studies. It may be that this represents active avoidance of minor, insignificant threats, such as words and pictures. It would clearly be adaptive to avoid the many distractions of minor threats and single out only serious threats for particular attention.

### **ACTIVITY 13.7**

---

Can you think of a situation where you paid attention to negative cues in your environment because of a fear that you have? If you haven’t noticed this negative bias in yourself, have you ever noticed another, non-negative, attentional bias towards features in the environment?

#### **COMMENT**

Anyone who is very afraid of spiders might recognize this characteristic in themselves. Almost invariably they will notice any spider in the surroundings well before their non-phobic companions. Most people have a bias to attend to things that match their special interests. Temporary biases are also common, and can, for instance, occur when you have acquired something new such as when you purchase a new car. For a while you may find yourself noticing many examples of this same model which previously you ignored. This might give you a feel for what it could be like for an anxious individual, although for them it is unpleasant items that just constantly catch their attention without any intention on their part.

---

Although we have mostly mentioned anxious patients so far, biases favouring attention to threat can also be found in high trait anxious participants, although less reliably. Moreover, these biases tend to be stronger when high state anxious mood and high trait anxiety occur together. Anxious patients tend to have relatively high levels of state anxiety much of the time so it is unsurprising that attentional biases are more robust in this group.

### **4.4 Semantic interpretation**

Semantic interpretation (see Chapter 6) is another cognitive process known to be influenced by emotion. If you see a word such as ‘batter’, do you think of pancakes or do you think of an assault on an innocent victim? It is surprising how many situations in life can be ambiguous and therefore open to biases of interpretation. In this section we shall consider interpretation of ambiguous linguistic information, but be aware that the same processes apply in many situations. Assessing the nature of the shadow in the path ahead on a dark night, or guessing the meaning of the probing look of the interviewer when you apply for a job are just two other such examples.

The earliest work on interpretation and emotion used **homophones**. These are words like ‘pane’ and ‘pain’ or ‘die’ and ‘dye’ which sound the same but have

different spellings associated with different meanings. Eysenck *et al.* (1987) asked both high and low trait anxious individuals to write down the homophones as they heard them. All the homophones had both a threatening/unpleasant and a non-threatening or neutral meaning. This simple technique revealed which interpretation had been made, by the spelling which participants chose. They found that the higher the participant's trait anxiety, the more threat spellings they produced. This indicated that trait anxiety was linked to a tendency to assume the negative interpretation of an ambiguous stimulus – i.e. an *interpretative bias*.

However, this method soon fell foul of criticism. For example, it is possible that participants were aware of and had access to both spellings, but just chose to write down the negative one. This matters because, if true it would mean that there was no bias in the actual *interpretation* of the words – both interpretations were made. Instead the bias would be at the stage of making the response, which then says little about the cognitive processing involved in making interpretations.

Later work used an alternative method to avoid this and other problems. For example, in their classic study Richards and French (1992) used **homographs** instead of homophones. These are words which have dual meanings, despite having the same spelling, such as ‘batter’, ‘punch’ and ‘stalk’. They used these words in a priming experiment involving a lexical decision task (a task described in Chapter 2, Section 1.3). This task involves simply identifying, as rapidly as possible, whether the second of two sequentially presented items is a real word or a meaningless letter string (a non-word). From the participant’s point of view the first item that appears is just to be ignored. However, this first word is actually a prime.

As described in Chapter 2, if the prime is related in meaning to the second word, the target, (e.g. cat–dog, nurse–doctor) then lexical decisions are expected to be speeded compared to primes and targets which bear no semantic relation (e.g. cat–doctor, nurse–dog).

We can use this logic to infer how participants interpreted the homograph primes. For example, if lexical decisions for trials like batter–assault were faster than for trials like batter–pancake, this would imply that the participant interpreted batter as ‘assault’ rather than ‘pancake’. The results of the Richards and French study, as well as other similar studies, suggest that high anxious participants show a negative bias in interpretation – that is, there is a greater priming effect for target words related to the negative meaning of the homograph than the neutral meaning. For non-anxious participants there is, once again, the familiar positive bias towards the more positive or non-threatening meaning. Further studies have extended this research by using ambiguous sentences or even passages of text, for example:

‘The doctor examined little Emily’s growth’ (her height or her tumour?)

‘The two men watched as the chest was opened’ (a gruesome operation or an exciting find?)

‘Your friend asks you to give a speech at her wedding reception. You prepare some remarks and when the time comes, get to your feet. As you speak, you notice some people in the audience start to laugh’ (appreciatively, or rudely?)

The concept of protective processing styles such as these has been described formally in a theory known as **attribution theory**. A common observation is that we attribute good things internally, as something within our control, whereas bad

things are attributed externally to others or to circumstances. This reflects a tendency to accept the credit for good outcomes and blame something or someone else for bad outcomes. For example, if you are late for an important meeting or fail your driving test you might say ‘I’m terribly sorry but the train times have changed and I couldn’t help being late’ or ‘I had such an unreasonable examiner’ or ‘My instructor gave me inadequate preparation’; if you are early or on time, or pass your test first time, you might well congratulate yourself for your efficient organization and planning, or excellent driving skills. You may have come across this described elsewhere as the **self-serving attribution bias**.

Although these self-serving biases might seem an irrational way of thinking, the evidence repeatedly supports their existence and, as with other positive biases, they may have protective properties. Moreover, in emotional disorders, particularly in depression or anxiety, we know that this self-serving bias can be lost or even reversed. Such people might think passing the driving test was just luck, or the examiner being lenient, whereas failing was yet more evidence of their own worthlessness and lack of skill. In some situations it can be shown that by lacking the positive bias the depressed person’s attribution of their own performance can be more accurate than for non-depressed controls, so-called ‘depressive realism’.

### **ACTIVITY 13.8**

---

When you are next chatting with family or friends, or watching conversations on the television, see if you can identify some of the attributions people make. Does this go along with the attribution theory? Do you notice any examples of the self-serving attribution bias?

---

It should be noted that, although the various positive biases that we have described are thought to be quite normal, and have protective qualities (such as helping to maintain good mood and a positive self-image), it is equally true that, taken to their limits, they would be maladaptive.

## **Summary of Section 4**

---

- State emotion refers to the feelings of the moment whereas traits refer to more enduring personality characteristics of an individual.
- The manifestation of emotion is distinct from the processing of emotional material. The former refers to feelings, behaviours and bodily responses. The latter refers to the emotional content of the external stimuli upon which the cognitive system acts.
- The field of cognition and emotion is primarily concerned with the conjunction between state or trait emotions and the processing of emotional material.
- Mood congruent memory (MCM) refers to enhanced memory for material that matches present mood. The phenomenon is particularly apparent in depression and may contribute to the clinical disorder.

- Mood dependent memory (MDM) occurs when recall is enhanced by a match between mood at the time of learning and mood at the time of testing. However the effect is not very robust. Bower's semantic network theory provides one explanation of MDM and MCM.
- Biases in attention and in semantic interpretation are associated with both trait and state anxiety, and again may contribute to chronic anxiety and clinical anxiety disorders.

## 5 Does cognition influence emotion?

### 5.1 A look at some historical answers

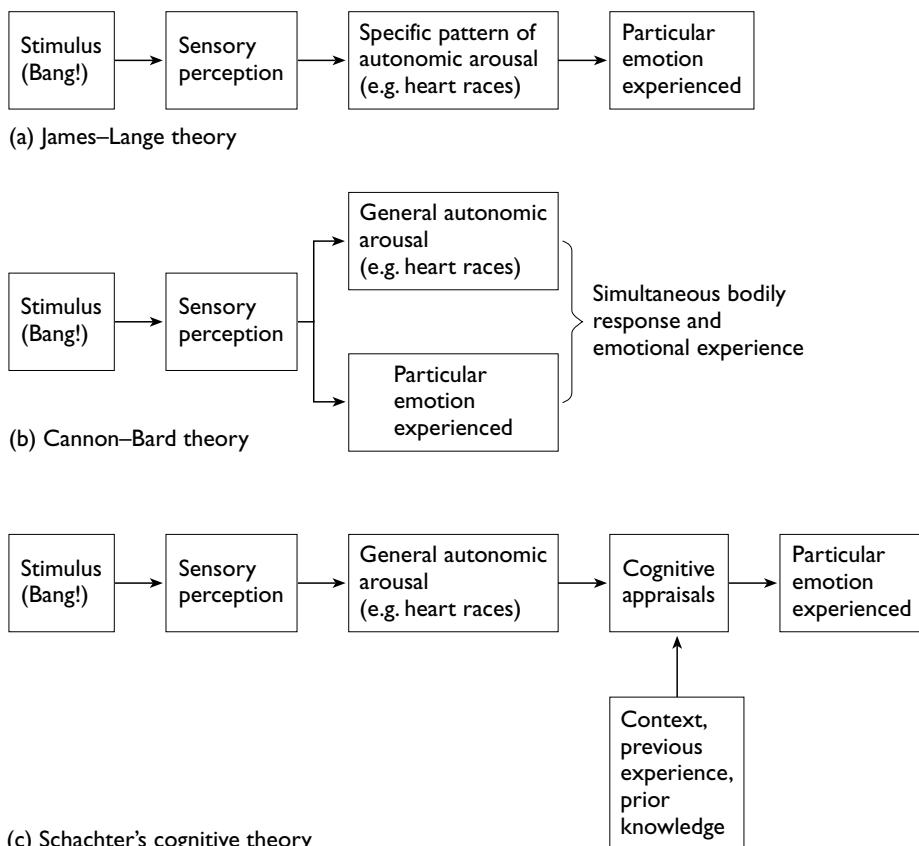
Do we laugh *because* we feel happy or is it the laughing itself that *makes* us feel happy? This question has been central to emotion research since its very beginnings back in the 1880s when William James (1843–1910), commonly regarded to be one of the founders of psychology, first considered it. Putting the question another way, is it our experience of the behavioural and bodily responses associated with emotion that make us subjectively feel that emotion? Or do those responses follow on from our subjective experience of emotion?

#### 5.1.1 James-Lange

James's answer to this question in the late nineteenth century was the counter-intuitive one. Namely, he argued that we feel fear *because* we run and we experience happiness as a *result* of laughing: the cognitive and experiential side of emotion was a slave to the physiology of emotion. Carl Lange took a very similar position, and so this view became known as the 'James–Lange' theory (see Figure 13.13(a)). Their observation was that behaviour, most especially in a frightening situation, was initiated too rapidly to have arisen from a *feeling* of fear that was subsequently translated into a conscious decision to act. Rather, they felt that behaviour preceded (conscious) cognition, and more precisely that the experience of emotion depended on the behaviour and bodily reaction that followed an event. More recent studies, such as those of LeDoux (1996) looking at the speed of the startle response to loud noise, indicate that these responses are initiated within a few milliseconds, well before conscious awareness has time to develop. One implication of this way of looking at things is that physiological responses and behaviours must be distinct and occur in unique constellations in order that different emotions actually occur and feel different. Love and fear feel different, according to James, because they result from different physiological signatures.

#### 5.1.2 Cannon-Bard

Walter Cannon and Philip Bard challenged this view in the 1920s precisely because they felt that physiological responses were pretty indistinguishable across most emotions, and indeed that similar physiological patterns (e.g. increased heart rate, sweating and inhibited ingestion) could arise from fever during illness. According to



**Figure 13.13** Comparing the theories

Source: based on Rosenzweig et al., 1999, Figure 15.1, p.412

them, what distinguishes one emotion from another, given this common physiology, is the pattern of cortical stimulation that arises. For Cannon and Bard, both the autonomic arousal and the subjective experience of a specific emotion could occur simultaneously and were instigated by the higher brain areas such as the cerebral cortex (see Figure 13.13(b)). Thus for Cannon–Bard you don't have to cry (or be suppressing the tendency to do so) to feel sad – there simply has to be an appropriate activation of the thalamus. For the James–Lange theory, preventing crying (and any urge to cry) would prevent sadness.

Both sets of theorists could provide evidence to support their view. The Cannon–Bard camp challenged the view that physiological responses were sufficiently unique to distinguish between the emotions (or indeed between emotion and other causes). They also reasoned that animals and humans with damaged spinal cords, preventing normal physiological responses, nevertheless responded emotionally. In support of the James–Lange theory, emotional response does seem to be blunted in those unfortunate enough to have suffered spinal injury preventing both physiological changes and overt behaviour in response to emotional situations, and this occurs in proportion to the loss of sensation. Emotional feeling is not,

however, absent. There must, therefore, be more to emotional feeling than interpreting the sensation of body movement and physiological responses. The debate about the specificity of physiological responses to each emotion, however, still continues. Ekman and colleagues have concluded that there are emotion-specific physiologies for at least anger, fear, disgust and perhaps sadness (Ekman *et al.*, 1983; Levenson *et al.*, 2001). For example, Ekman claims that skin conductance (discussed in Box 13.1) is higher during sadness than the other emotions, and heart rate decelerates during disgust.

However, John Cacioppo, a renowned psychophysiologist, disagrees. Having reviewed all the available data he concluded (Cacioppo *et al.*, 2000) that the evidence for emotion-specific physiology is far more uncertain and suggests that discrete emotions cannot be differentiated by autonomic activity alone. However, he does agree that there may be a consistent distinction between the positive and negative emotions in general. He proposes that negative emotions are associated with greater motivational output than positive and therefore show generally greater levels of autonomic activation.

### 5.1.3 Schachter–Singer

When cognitive psychology began to take hold in the 1960s, Stanley Schachter and Jerome Singer proposed an alternative to both the earlier James–Lange and Cannon–Bard views. Like James they held that physiological mechanisms were crucial, but like Cannon they also believed that these responses were non-specific and could not distinguish the different emotions. Instead they thought differentiation was achieved by the individual's particular interpretations or attributions about why those bodily responses were occurring. These ambiguous messages from the body were interpreted by taking into account things like the situational context, previous experience of when certain emotions occur, expectations and intellectual knowledge of the world. Physiological arousal may be responsible for feeling emotion but cognitive interpretations, or **cognitive appraisals**, were what distinguished one emotion from another (see Figure 13.13(c)).

The Schachter–Singer theory predicts that it should be possible to change our experience of emotion by changing the cognitive appraisals we make, even if the physiological signs remain the same. To test this they performed a now famous experiment (Schachter and Singer, 1962). Participants were injected with epinephrine (adrenalin), which is a hormone that stimulates activity in the sympathetic nervous system (discussed in Section 1.1.2). Some participants were told that there would be no effect of this injection while others were told that it would make their heart race. The latter group did not report any emotional experience, while the former group did. What does this result say about the James–Lange theory of emotion? The fact that those expecting no effect of injection did actually experience emotion is consistent with James–Lange – the physiology directly led to an emotional experience. However, the other result is inconsistent with that theory. Those able to attribute the bodily sensations to the injection failed to experience emotion. The James–Lange theory makes no allowance for such cognitions to influence emotional experience in this way. It would predict that, despite this knowledge, the physiological arousal should directly give rise to an emotional experience.

In a further aspect of the above experiment, participants were put in a room with a ‘stooge’ who was party to the experiment and who acted either in an extremely happy or very angry manner. The behaviour of the stooge directly influenced the feelings of those participants who reported experiencing emotion (i.e. those with no foreknowledge of the effects of the injection). Those with the happy stooge reported experiencing happiness, whereas those with the angry stooge reported anger. This demonstrates a strong influence of context on the specific emotion experienced. How does this second finding fit with the James–Lange theory? It is clearly inconsistent with that theory because the same physiological reactions were being experienced differently according to context.

Schachter and Singer used the results of this experiment to support their theory. The non-specific physiological arousal interacted with the social and physical context that participants experienced to determine the precise emotion that was felt. They had succeeded in showing that identical physiological states could be subjectively experienced as different emotions according to how the individual appraised their circumstances. Similarly, subjective emotional experience could be eliminated by telling participants the true source of their bodily sensations. This convincingly showed that cognitive attributions were crucial in whether or not emotions were experienced in their fullest sense.

The Schachter–Singer theory has been criticized. For example, we now know, because we can measure them, that physiological responses associated with different emotions are not in fact identical (and their results have not always been replicated; e.g. Reisenzein, 1983). However, the lasting contribution of this theory was the notion of cognitive appraisal being critically involved in the generation of emotion. The acceptance of this possibility spawned a whole generation of appraisal theorists.

#### **5.1.4 Appraisal theories today**

The central idea of appraisal theories is that emotions are elicited and differentiated on the basis of the individual’s subjective evaluation of the external situation or event combined with their own physiological state. Although this answers the question of how we differentiate between emotions, it simultaneously raises others. How are these appraisals made – using what yardstick, measuring tool or criteria?

Each different appraisal theory tends to suggest different dimensions that we use when making appraisals. For example, Klaus Scherer proposes specific fixed sets of criteria that we supposedly apply to any situation that comes our way. For example, one criterion is the intrinsic qualities of the event such as how novel or agreeable it is. Another is how significant the event is in terms of our own personal goals or needs. Clearly an event which is neutral in terms of our goals or needs, is unlikely to generate emotion. To get an idea of how his criteria can distinguish one emotion from another, look at Table 13.3 overleaf.

**Table 13.3** Scherer's appraisal criteria and profiles for different emotions

<b>Stimulus evaluation checks</b>	<b>Anger/rage</b>	<b>Fear/panic</b>	<b>Sadness</b>
<b>Novelty</b>			
● Suddenness	High	High	Low
● Familiarity	Low	Various	Low
● Predictability	Low	Low	Various
<b>Intrinsic pleasantness</b>	Various	Various	Various
<b>Goal significance</b>			
● Concern relevance	Order	Body	Various
● Outcome probability	Very high	High	Very high
● Expectation	Dissonant	Dissonant	Various
● Conduciveness	Obstruct	Obstruct	Obstruct
● Urgency	High	Very high	Low
<b>Coping potential</b>			
● Cause: agent	Other	Other/nature	Various
● Cause: motive	Intent	Various	Chance/neg
● Control	High	Various	Very low
● Power	High	Very low	Very low
● Adjustment	High	Low	Medium
<b>Comparability with standards</b>			
● External	Low	Various	Various
● Internal	Low	Various	Various

Note: 'Various' = different appraisal results are compatible with the respective emotion.

Source: Dalgleish and Power, 1999, Table 30.2, p.639

For example, in the table, items under the fear/panic category suggest that this emotion results from an event (stimulus) that is judged to be 'high' on novelty/suddenness, 'low' on novelty/predictability; it is of concern to the body's status (goal significance/concern relevance); urgency is high, coping potential/power is very low, and so on. For a number of the appraisals involving fear/panic there can be various options; for instance, under novelty/familiarity it is possible to be afraid of something either familiar or unfamiliar so that this is not a defining feature for that emotion. To give a concrete example, imagine the consequence of a spider emerging from under the sofa for a spider phobic: this is a sudden event; but spiders are familiar (although disliked); they behave unpredictably; their intrinsic pleasantness is 'various', that is, some people find them pleasant (although a spider phobic certainly would not); the ability to cope is perceived as low, and so on. You may feel that it would be impossible to evaluate an event on all these various things *before* feeling the emotion, there surely would not be enough time? This is a fair criticism and one that has been made by opponents of appraisal theory. However, the counter-

argument to this is that appraisals do not have to be conscious serial processes; they may well occur in parallel, automatically.

The evidence in support of appraisal theory relies entirely on subjective self-report and for this reason these theories have been heavily criticized. Typically, participants are either asked to remember personal events, or are exposed to experimental manipulations designed to induce an emotion. They are then asked to report, either verbally or using questionnaires, the types of appraisals they engaged in. You may have spotted another problem with this. By definition, participants will not be able to report on any appraisals made unconsciously, and this is a second major criticism of the evidence for appraisal theory. As yet, appraisal theorists have only been able to counter this by stating that no alternative to self-report exists. It will be interesting to see whether appraisal theorists will be able to find ways of accessing automatic evaluations (perhaps using brain-imaging technologies).

## 5.2 A clash of minds: the cognition/emotion debate

No discussion of cognition and emotion would be complete without considering one famous example of the different approaches psychologists can take. The two main protagonists in the debate were Richard Lazarus and Robert Zajonc (pronounced zy-unce, to rhyme with once).

### 5.2.1 Zajonc's view

Zajonc disagreed with appraisal theory's contention that emotions are produced by cognitive processes. He challenged the appraisal theorists directly (Zajonc, 1980) making two key assertions:

- 1 Appraisal is not necessary for emotion to be experienced. Emotions could arise directly without the need for cognitions at all. This is similar to the James–Lange idea in that cognition plays no part in the process of eliciting emotion.
- 2 The experience of emotion always precedes one's cognitive processing of that emotion. This stronger claim adds to the first by saying that not only is appraisal not necessary, in fact it never occurs before the emotional experience itself. This question of whether emotion precedes cognition, or the other way round, is known as the **primacy debate**.

The following quotation summarizes Zajonc's position very well: 'Affect [meaning mood or state emotion] and cognition are separate and partially independent systems and ... although they ordinarily function conjointly, affect could be generated without a prior cognitive process' (Zajonc, 1984, p.117).

How is this issue of primacy different from William James's question? Do we laugh because we're happy or are we happy because we laugh? James was concerned with the relationship between the conscious feeling of experiencing an emotion and the physiological and behavioural expression of that emotion (see Sections 1.1.1 and 1.1.2). The concept of cognitive appraisal had not yet been articulated. The primacy debate contrasts the cognitive appraisal of an emotion with all its other aspects (feeling, physiology and behaviour).

In support of his argument Zajonc described an experiment which used the famous **mere exposure** effect. Mere exposure refers to the finding that people tend to

prefer items to which they have previously been exposed over comparable novel ones. Simple familiarity with something creates a preference for that item. This is presumably one reason for the success of the advertising industry. Zajonc took the mere exposure method and adapted it so that items were presented **subliminally** (below the level of conscious awareness) while participants were engaged in another, primary task. His results revealed that while participants showed no recognition of the subliminal items, they nevertheless gave them higher preference ratings than novel items!

Zajonc argued that these results showed that cognition was not necessary in order to have affective experience. He was assuming, first, that stimuli were not being processed ‘cognitively’ because they were presented subliminally. Second, he was assuming that preference ratings were tantamount to emotional experience. Both these assumptions have since been challenged. Today the details of nonconscious processing (outside awareness) are controversial, but few would challenge its existence (see Chapter 15). Certainly, it is unlikely that many psychologists now accept Zajonc’s implicit assumption that all cognitive processing must be conscious. Likewise equating preference judgements with affect or emotion is probably a step too far. Surely only very limited emotion is involved in rating how much you like something that has no particular meaning or relevance to you?

### 5.2.2 Lazarus’s view

Richard Lazarus, on the other hand, argued that cognitive appraisal was essential for the experience of emotion: ‘Cognitive appraisal (of meaning or significance) underlies and is an integral feature of all emotional states’ (Lazarus, 1982, p.1021).

In support of his position he undertook several studies. Typically, emotions would be elicited by showing participants anxiety-provoking films. For example, one was a Stone Age circumcision ritual (another showed someone involved in a gruesome industrial accident – it is unlikely that this type of material would obtain ethical approval for use today!). Cognitive appraisal was manipulated by playing one of two soundtracks while participants watched the films. A ‘denial’ soundtrack included statements indicating that one was a safety film, the people in the films were actors and the ritual in the film was not actually painful. An ‘intellectualization’ soundtrack emphasized an anthropological perspective and advocated, for example, considering the ritual as a strange native custom. A control condition had no soundtrack. Physiological measures such as GSR (galvanic skin response) and heart rate were taken throughout viewing and suggested that the appraisals produced by the soundtracks did indeed reduce emotional responses significantly compared with the control condition. Although impressive, these results did *not* prove that cognition necessarily precedes affect, but rather that cognitive appraisal *can* convincingly alter emotional response.

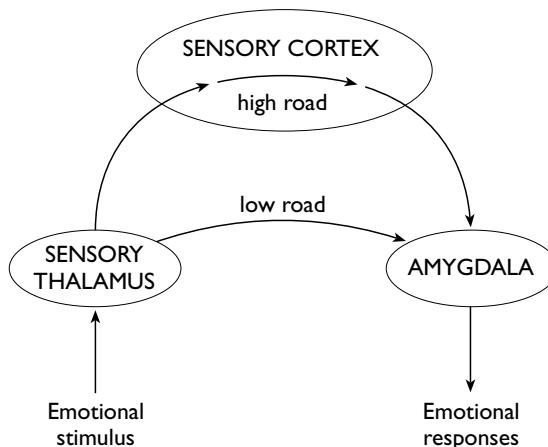
### 5.2.3 A resolution?

Despite the ferocity of their debate about primacy, neither protagonist marshalled sufficient evidence to win the argument. Rather a resolution was reached by both identifying their positions more clearly. Zajonc acknowledged the view that the existence of nonconscious appraisal was a key question, and Lazarus conceded that although appraisal might influence emotion this did not mean it was an essential

component. Both agreed that, as Zajonc puts it: ‘It is a critical question for cognitive theory and for theories of emotion to determine just what is the minimal information process that is required for emotion. Can untransformed pure sensory input directly generate emotional reactions?’ (Zajonc, 1984).

Interestingly, more recent work by Joseph LeDoux (LeDoux, 1989; LeDoux, 1996) has thrown further light on the issue of primacy, suggesting that Zajonc may be right after all. These studies used **lesioned** animals in which specific neural pathways within the animals’ brains were deliberately severed by the experimenter. Doing this allows an experimenter to deduce the function of the damaged pathways or regions by giving the animal various tasks to perform and establishing which of these are impaired. You may wish to think about the ethical issues such procedures raise, though we do not have room to consider them here.

Using a variety of tasks manipulating emotions, especially fear, LeDoux has shown that certain brain structures such as the thalamus and the amygdala play different roles in the generation of emotion (see colour Plate 9). Anatomical work has shown that these areas are connected via two routes, as you will see from Figure 13.14. The ‘lower’ route – so-called because only *evolutionarily old* structures are involved – takes sensory information from the primary sensory areas (the regions of the brain where sensory information arrives first) to the thalamus and then directly to the amygdala. This route bypasses the higher brain structures in the cortex and provides a fast thalamo-amygdala connection involving only one **synapse** (a relay junction between one nerve cell and the next). The ‘higher’ route – so-called because the *evolutionarily newer* areas such as the cortex are involved – relays information through a more complex route from the thalamus via the sensory cortex to the amygdala.



**Figure 13.14** The low road and the high road to the amygdala

Source: LeDoux, 1996, Figure 6.13, p.164

LeDoux has shown that learning about new fearful situations or altering existing knowledge about fear requires the higher thalamo-cortical-amygdala route to be intact and functioning. However, once fear responses are well learned, lesions to this higher route do not diminish the response. In such cases the lower thalamo-amygdala

route is sufficient. He suggests that the lower route is ‘quick and dirty’, providing a means of rapid identification of, and initiation of responses to, emotionally significant stimuli without the need for time-consuming higher processing. The higher route, on the other hand, is needed to learn new associations or re-learn or extinguish old ones. It is vital in the initial stages of learning about novel fear stimuli, or in the modification of original learning experiences, but when this learning has become automated, the lower, more rapid route seems capable of taking over.

You can see then that this quick and dirty route could be a neuroanatomical substrate (physical basis) for Zajonc’s idea of the direct elicitation of emotion without the need for cognition. Maybe the primacy of emotion over cognition was partially right after all. However, the role of the higher route also seems to map onto Lazarus’s idea that cognitive processing can precede, or at least influence, emotion. LeDoux suggests that the higher cortical route is necessary to override the quick and dirty route in certain situations – perhaps where the threat turns out not to be so bad, once cognitively processed more fully, or where past knowledge and experience suggest that an emotional reaction would be inappropriate. As is often the case in these debates, both the positions of Lazarus and Zajonc may turn out to be partially correct.

## Summary of Section 5

---

- A question that has always concerned psychologists is whether emotions arise before or after the cognitive processing of the stimuli and situation that elicits them. This is sometimes known as the primacy debate.
- Zajonc argued that emotions do not require prior cognition, whereas Lazarus maintained that cognitive appraisal was an integral part of the production of emotion.
- Recent neuroanatomical and lesion studies suggest that both may have been right after all. LeDoux argues for a ‘quick and dirty’ neural pathway enabling immediate response to potentially emotional stimuli, without the need for prior cognitive processing. He also proposes a slower cortical route enabling learning and modulation of emotional responses according to the outcome of cognitive appraisals.

## 6 General summary

In this chapter we began by noting that emotions comprise at least three components: feelings that can only be reported through introspection; behaviours that can be observed; and bodily responses, some of which can be precisely measured using psychophysiological techniques. We discussed two different approaches to classifying emotions: that of dividing emotional experience up into separate ‘basic’ emotions; and that of using two or more dimensions to describe a continuum of experience. We briefly explored what some of the functions of emotion might be,

such as their use for survival, to enhance performance or as information signals to tell us how to behave (the somatic marker hypothesis).

For the rest of the chapter we considered the interaction between emotions and cognition, first by looking at how emotional material is processed differently from non-emotional material. In memory, attention and semantic interpretation we saw how biases in processing usually operate to favour the processing of emotionally significant information. This is a particularly important topic because psychologists believe such biases contribute to emotional disorders such as anxiety and depression. Second, we saw how cognitive appraisals can influence the experience of emotion, and several theoretical variations on this basic idea were described.

Emotion and its interaction with cognition is becoming an increasingly popular area of psychology, helped in part by the availability of new brain-imaging technologies. Armed with the knowledge you have gleaned from this chapter you will be in an excellent position to follow the progress of this exciting area of psychology.

## Further reading

- Dalgleish, T. and Power, M.J. (1999) *Handbook of Cognition and Emotion*, Chichester, Wiley.
- LeDoux, J.E. (1996) *The Emotional Brain*, New York, Simon and Schuster.
- Williams, J.M.G., Watts, F.N., MacLeod, C. and Mathews, A. (1997) *Cognitive Psychology and Emotional Disorders* (2nd edn), Chichester, Wiley.

## References

- Bower, G.H. (1981) 'Mood and memory', *American Psychologist*, vol.36, pp.129–48.
- Cacioppo, J.T., Bernston, G.G., Larsen, J.T., Poehlmann, K.M. and Ito, T.A. (2000) 'The psychophysiology of emotion', in Lewis, M. and Haviland-Jones, J.M. (eds) *Handbook of Emotions* (2nd edn), London, The Guilford Press.
- Calder, A.J., Lawrence, A.D. and Young, A.W. (2001) 'Neuropsychology of fear and loathing', *Nature Reviews Neuroscience*, vol.2, no.5, pp.352–63.
- Calder, A.J., Keane, J., Manes, F., Antoun, N. and Young, A.W. (2000) 'Impaired recognition and experience of disgust following brain injury', *Nature Neuroscience*, vol.3, pp.1077–8.
- Dalgleish, T. and Power, M.J. (1999) *Handbook of Cognition and Emotion*, Chichester, Wiley.
- Damasio, A.R. (1996) 'The somatic marker hypothesis and the possible functions of the prefrontal cortex', *Philosophical Transactions of the Royal Society of London*, Series B, vol.351, pp.1413–20.
- Darwin, C. (1998, first published 1872) *The Expression of the Emotions in Man and Animals* (3rd edn), London, Harper Collins.
- Dawson, M.E., Schell, A.M. and Bohmelt, A.H. (eds) (1999) *Startle Modification*, Cambridge, Cambridge University Press.

- Eibl Eibesfeldt, I. (1988) 'Social interactions in an ethological, cross-cultural perspective', in Poyatos, F. (ed.) *Cross Cultural Perspectives in Nonverbal Communication*, Kirkland, WA, Hogrefe & Huber Publishers.
- Eich, E. and Metcalfe, J. (1989) 'Mood dependent memory for internal versus external events', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.15, pp.443–55.
- Ekman, P. (1973) 'Universal facial expressions in emotion', *Studia Psychologica*, vol.15, pp.140–7.
- Ekman, P. (1999) 'Facial expressions', in Dalgleish, T. and Power, M.J. (eds) *Handbook of Cognition and Emotion*, New York, John Wiley & Sons Ltd.
- Ekman, P. and Friesen, W.V. (2003) *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*, Cambridge, MA, Malor Books.
- Ekman, P. and Oster, H. (1979) 'Facial expressions of emotion', *Annual Review of Psychology*, vol.30, pp.527–54.
- Ekman, P., Levenson, R.W. and Friesen, W. (1983) 'Autonomic nervous system activity distinguishes among emotions', *Science*, vol.221, pp.1208–10.
- Ekman, P., Sorenson, E.R. and Friesen, W.V. (1969) 'Pan-cultural elements in facial displays of emotion', *Science*, vol.164, pp.86–8.
- Eysenck, M. and Keane, M.T. (1996) *Cognitive Psychology: A Student's Handbook* (3rd edn), Hove, Psychology Press.
- Eysenck, M.W., MacLeod, C. and Mathews, A.M. (1987) 'Cognitive functioning in anxiety', *Psychological Research*, vol.49, pp.189–95.
- Frijda, N.H. (2001) 'The self and emotions', in Bosma, H.A. (ed.) *Identity and Emotion: Development Through Self Organization, Studies in Emotion and Social Interaction*, New York, Cambridge University Press.
- Frijda, N.H. and Tcherkassof, A. (1997) 'Facial expressions as modes of action readiness', in Russell, J.A. and Fernández Dols, J.M. (eds) *The Psychology of Facial Expression, Studies in Emotion and Social Interaction, Second Series*, New York, Cambridge University Press.
- Lazarus, R.S. (1982) 'Thoughts on the relations between emotion and cognition', *American Psychologist*, vol.37, pp.1019–24.
- LeDoux, J.E. (1989) 'Cognitive-emotional interactions in the brain', *Cognition and Emotion*, vol.3, pp.267–89.
- LeDoux, J.E. (1996) *The Emotional Brain*, New York, Simon and Schuster.
- Levenson, R.W., Cacioppo, J.T., Davidson, R.J., Lang, P., Ohman, A. and Stemmler, G. (2001) 'Psychophysiology of emotion: a decade and a half later', *Psychophysiology*, vol.38, Supplement S4.
- MacLeod, C., Mathews, A. and Tata, P. (1986) 'Attentional bias in emotional disorders', *Journal of Abnormal Psychology*, vol.95, no.1, pp.15–20.
- Mathews, A. (1990) 'Why worry? The cognitive function of anxiety', *Behaviour Research and Therapy*, vol.28, no.6, pp.455–68.
- Mathews, A. and MacLeod, C. (2002) 'Induced processing biases have causal effects on anxiety', *Cognition and Emotion*, vol.16, no.2, pp.331–54.

- Matt, G.E., Vacquez, C. and Campbell, W.K. (1992) 'Mood-congruent recall of affectively toned stimuli: a meta-analytic review', *Clinical Psychological Review*, vol.12, pp.227–55.
- Medicus, G., Schleidt, M. and Eibl Eibesfeldt, I. (1994) 'Universelle Zeitkonstante bei Bewegungen taubblinder Kinder' ('Universal time constancy in movements of deaf-blind children'), *Nervenarzt*, vol.65, no.9, pp.598–601.
- Oatley, K. and Jenkins, J.M. (1996) *Understanding Emotions*, Oxford, Blackwell Publishers Ltd.
- Oatley, K. and Johnson-Laird, P.N. (1987) 'Towards a cognitive theory of emotions', *Cognition and Emotion*, vol.1, pp.29–50.
- Ortony, A. and Turner, T.J. (1990) 'What's basic about basic emotions?', *Psychological Review*, vol.97, no.3, pp.315–31.
- Panksepp, J. (1989) 'The neurobiology of emotions: of animal brains and human feelings', in Wagner, H. and Manstead, A. (eds) *Handbook of Social Psychophysiology*, Wiley Handbooks of Psychophysiology, Oxford, John Wiley and Sons.
- Panksepp, J., Sacks, D.S., Crepeau, L.J. and Abbott, B.B. (1991) 'The psycho- and neurobiology of fear systems in the brain', in Denny, M.R. (ed.) *Fear, Avoidance, and Phobias: A Fundamental Analysis*, Hillsdale, NJ, Lawrence Erlbaum Associates Inc.
- Plutchik, R. and Landau, H. (1973) 'Perceived dominance and emotional states in small groups', *Psychotherapy: Theory, Research and Practice*, vol.10, pp.341–2.
- Power, M.J. and Dalglish, T. (1997) *Cognition and Emotion: From Order to Disorder*, Hove, Psychology Press.
- Reber, A.S. (1995) *The Penguin Dictionary of Psychology* (2nd edn), Harmondsworth, Penguin.
- Reisenzein, R. (1983) 'The Schachter theory of emotions: two decades later', *Psychological Bulletin*, vol.94, pp.239–64.
- Richards, A. and French, C.C. (1992) 'An anxiety-related bias in semantic activation when processing threat/neutral homographs', *Quarterly Journal of Experimental Psychology*, vol.45, pp.503–25.
- Rosenzweig, M.R., Leiman, A.L. and Breedlove, S.M. (1999) *Biological Psychology: An Introduction to Behavioural, Cognitive and Clinical Neuroscience* (2nd edn), Sunderland, MA, Sinauer Associates Inc.
- Schachter, S. and Singer, J. (1962) 'Cognitive, social and physiological determinants of emotional state', *Psychological Review*, vol.69, pp.379–99.
- Scherer, K.R. and Wallbott, H.G. (1994a) 'Evidence for universality and cultural variation of differential emotion response patterning', *Journal of Personality and Social Psychology*, vol.66, no.2, pp.310–28.
- Scherer, K.R. and Wallbott, H.G. (1994b) 'Evidence for universality and cultural variation of differential emotion response patterning: correction', *Journal of Personality and Social Psychology*, vol.67, no.1, p.55.

- Segal, Z.V., Williams, J.M.G. and Teasdale, J.D. (2002) *Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse*, New York, Guilford Press.
- Stroop, J.R. (1935) 'Studies of interference in serial verbal reactions', *Journal of Experimental Psychology*, vol.18, pp.643–62.
- Teasdale, J.D. (1988) 'Cognitive vulnerability to persistent depression', *Cognition and Emotion*, vol.2, no.3, pp.247–74.
- Wallbott, H.G. and Scherer, K.R. (1988, first published 1986) 'How universal and specific is emotional experience? Evidence from 27 countries on five continents', in Scherer, K.R. (ed.) *Facets of Emotion: Recent Research*, Hillsdale, NJ, Lawrence Erlbaum Associates Inc. (First published in *Social Science Information (Sur les sciences sociales)*, vol.25, no.4, pp.763–95.)
- Yerkes, R.M. and Dodson, J.D. (1908) 'The relation of strength of stimulus to rapidity of habit-formation', *Journal of Comparative Neurology and Psychology*, vol.18, pp.459–82.
- Zajonc, R.B. (1980) 'Feeling and thinking: preferences need no inferences', *American Psychologist*, vol.35, pp.151–75.
- Zajonc, R.B. (1984) 'On the primacy of affect', *American Psychologist*, vol.39, pp.117–23.

# Autobiographical memory and the working self

## Chapter 14

Martin A. Conway and Emily A. Holmes

---

### 1 What are autobiographical memories?

Consider the following memories:

- 1 A memory freely recalled by a 54-year-old recalling memories from any point in his life:

I remember a bright sunny morning walking down a hill near our house. I had on a red jacket, red shirt, blue jeans, and brown suede boots. I was seventeen. I was going into town and I felt great ... it was a feeling of being sort of utterly calm, utterly well, a feeling of expectancy: interesting things were about to happen. It was a feeling I don't think I have had in such a 'pure' form since.

*(Taken from an unpublished study by Martin A. Conway)*

- 2 A response made by a person asked to recall a memory to the (cue) word 'ship':

We were going on holiday to France. I remember that we stayed at a boarding house in Dover and went down to the ferry very early the following morning. My brother and I were wildly excited it was the first time we had been abroad and the first time we had been on a ship of any sorts. I have a vivid memory of looking back at the White Cliffs as the boat pulled out of the harbour – they seemed immensely tall.

*(Conway, 1996)*

- 3 A memory recalled when reading about 'flashbulb' memories – vivid memories of one's personal circumstances when learning an item of news (Brown and Kulik, 1977):

My own memory for the declaration of the Second World War, from September 1939, occurred when I was aged 6 years and 6 months. I have a clear image of my father standing on the rockery of the front garden of our house waving a bamboo garden stake like a pendulum in time with the clock chimes heard on the radio which heralded the announcement. More hazily, I have an impression that neighbours were also out in the adjoining

gardens listening to the radio and, although my father was fooling around, the feeling of the memory is one of deep foreboding and anxiety.

*(Gillian Cohen, personal communication, 1994, see Conway and Pleydell-Pearce, 2000)*

- 4 A memory reported by David Pillemer in a study of what have been termed ‘self-defining’ memories (Singer and Salovey, 1993; see too Pillemer, 1998):

I remember sitting in ‘X’’s class on the day that a midterm ... was handed back. I was a freshman and felt that I was in over my head. The professor gave a stern lecture on the values of good writing before she handed back the papers. As she reproached us, my terror grew because her comments seemed to be personally directed at me. I was from a small town, did not have the same background as anyone in my class, and had immediately felt my inadequacies when class began in September. Then she said ‘But ‘Y’ has answered the questions well and has an unusual lyrical and personal style that enhanced her answer.’ I couldn’t believe that she was talking about my paper, but she was. I can still envision that dimly lit little room in the bottom of Z and smell its peculiar musty odour. I can still picture her stern but kind face and feel the relief and pride I felt at that moment.

*(Pillemer; 1998)*

- 5 A memory for a traumatic experience reported by a person suffering from post-traumatic stress disorder (PTSD):

A man who drove cars for a living was involved in a road traffic accident. He was a back seat passenger in a car when it was in a high speed collision with another vehicle; activation of the air bags in the front of the car produced a cloud of powder, which he thought at the time was smoke. At the time he could smell petrol and thought the car might ignite and remembered thinking ‘I will be burned alive.’ His wife was unconscious after the impact and he thought that she had died. He remembered thinking to himself ‘what am I going to do now?’ as he thought about his future alone without his wife. He had been experiencing terrific guilt about this as it suggests to him that he is a selfish person. In addition, he was an experienced driver and anticipated the crash, but did not cry out. He felt that he could have averted the crash if he had done this. He experienced intrusive thoughts, such as ‘I should have shouted’ (to warn the driver) and he relived the feeling he felt when he thought his wife had died, which he believed to be his fault because he had not shouted out.

*(Conway et al., 2004, see also Ehlers et al., 2004)*

Autobiographical memories like these, from the mundane to the profound, help form the self, they provide a personal historical context or personal biography for who we are now; they are in essence the ‘database’ of the self (Conway and Pleydell-Pearce,

2000; McAdams, 2001, see too Hollway and Jefferson, 2000). They help us integrate with each other, with the history of our times, and give a continuity to experience that would not otherwise be possible. Such a central form of cognition is, much as one might expect, highly complex and engages processes in many different parts of the brain. Because of this, autobiographical memory is highly susceptible to changes in brain function and is easily disrupted by brain injury (the experience of trauma) and by psychiatric illnesses. Complexity is also present in the nature of those memories that are freely retrieved and those that are recalled to cues, i.e. memory 2 above, and this is particularly evident in the distribution of memories across the lifespan. You might have already noticed in the example memories listed above that several date to when the rememberers would have been in their late 'teens and early twenties (memories 1, 2, and 4). This seems to be a time when particularly enduring memories are formed and memories from this period remain highly accessible, in contrast to memories from childhood and infancy, which are difficult to access. Indeed, in the example memories above none date to when the rememberers were five years or younger. In Section 2 we first consider the accessibility of memories across the lifespan. Section 3 concentrates on the psychological nature of autobiographical memories, their representation in long-term memory and their relation to the self. In Section 4 we review findings on disruptions of autobiographical memory following brain injury and the experience of trauma.

## 2 Autobiographical memory across the lifespan

### ACTIVITY 14.1

---

#### The lifespan retrieval curve

We are going to do an autobiographical memory retrieval experiment and you will need the following equipment: a pen, a stack of plain paper (say 20 to 40 sheets of A4 cut in half) and a watch (a stopwatch would be best but it is not essential).

You will need a quiet room to work in for about an hour. When you are ready read the instructions and start immediately.

##### *Instructions*

- 1 In the next 10 minutes recall as many memories as you can. The memories should be specific and detailed as in the examples at the start of the chapter. Try to sample from across your life and avoid recalling memories all from one period (for example, a recent holiday). No memories from the past 12 months are allowed.
- 2 Each time you recall a memory write down a short title on a piece of paper. The title should be designed so that if you read it again you would know exactly what you recalled. **IMPORTANT:** turn the sheet face down and do not look at the title again during this recall phase.

- 3 When the 10 minutes is up STOP.
- 4 DO NOT READ FURTHER UNTIL YOU HAVE RECALLED YOUR MEMORIES. THEN RETURN TO THE INSTRUCTIONS BELOW.

Now go back through and date each of the memories by recording how old you were, in months, when the recalled event took place (Age at Encoding or AaE). If you really want to simulate an autobiographical memory experiment you could also rate each of the memories on the rating scales used in Activity 14.2, ahead.

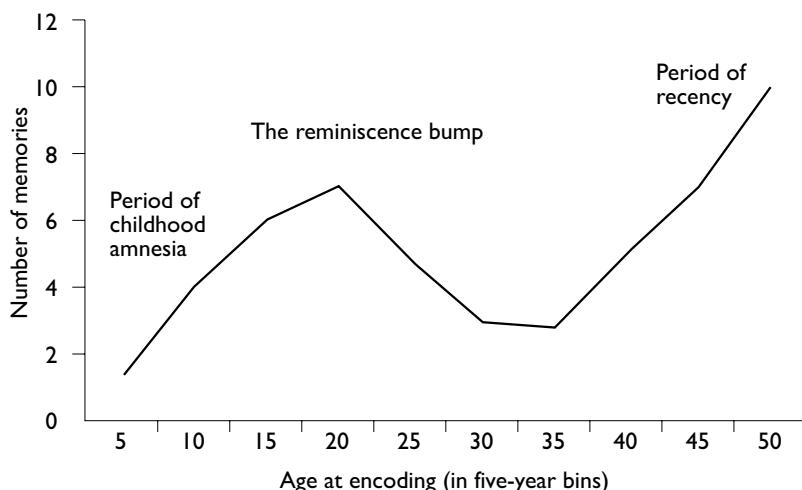
Now we want to plot AaE. To do this, have a scale ranging from '0' (birth) to your actual age now. Then divide the scale into five-year time bins (any size of 'bin' will do and I have chosen five year 'bins' or periods of time simply because this is often used in published reports). This AaE scale will form the 'X' axis running along the bottom of the graph. The 'Y' axis will simply be a count of the number of memories falling in each five-year time bin and will run from 0 to about 10 (it is unlikely that you will have more than 10 memories falling in any one time bin but if you do, increase the 'Y' axis scale to, say, 15 or 20, or whatever number best suits your data). For each bin in which memories occur mark an 'X' to indicate how many memories fall in that time bin. For example, maybe six memories date to the period when you were 20 to 25 years of age. The 'X' marked in this bin will then map on to '6' on the 'Y' axis. Next join up the 'X's and compare your lifespan retrieval curve to Figure 14.1.

---

Autobiographical memories are complex mental constructions that take time to bring to mind and once in mind have to be effortfully maintained. Although, of course, in abnormal remembering, such as occurs in PTSD (post-traumatic stress disorder, see Section 4), exactly the reverse may occur and some details of a trauma may be spontaneously and intrusively recalled and prove difficult to keep out of mind (such as the experience of guilt in memory 5, at the beginning of this chapter). Clearly, some set of central or executive processes must operate to construct memories appropriately, to keep irrelevant knowledge out of mind where it would intrude and take up resources needed for other tasks, and to ensure that what is recalled is relevant to the task or goal currently active. We have found it useful to postulate a structure we call the **working self** (Conway and Pleydell-Pearce, 2000). The choice of name is deliberate and it is intended to make an explicit connection to the concept of 'working memory' (Baddeley, 1986, 2000; see Chapter 9). The working self is conceived as a hierarchy of currently active goals (**goal hierarchy**) and self-conceptions through which current experience is encoded and in which memories are constructed. Because of this, we believe that the self has a profound influence on the accessibility of autobiographical knowledge and therefore upon the process of memory construction. This influence may extend across the lifespan, so that periods of change and development of the self, which contain self-defining memories that are crucial to the working self, may be particularly marked in autobiographical memory. The distinguishing aspect of memories and knowledge from these times may lie in their raised accessibility relative to other more dormant periods: in other words memories from these periods readily come to mind.

The working self – goal hierarchy and self-conceptions – probably first emerges in some more or less coherent form as the infant develops the ability for objective

and subjective self-awareness, i.e. conceptions of 'I' and 'Me', in its second year (Howe and Courage, 1997). Certainly children as young as 30 months have detailed autobiographical memories (Fivush *et al.*, 1996) although these typically are not accessible in adulthood. Undoubtedly the working self and its relation to autobiographical memory changes over the course of childhood and perhaps only stabilizes into an enduring form in late adolescence and early adulthood (Erikson and Erikson, 1982/1997). These periods of development of the self are reflected in the **lifespan retrieval curve**, which is observed when older adults (about 35 years and older) recall autobiographical memories in free recall or in a variety of cued recall conditions (Franklin and Holding, 1977; Fitzgerald and Lawrence, 1984; Rubin *et al.*, 1986; Rubin *et al.*, 1998). Memories are plotted in terms of age at encoding of the remembered experiences, and the resulting lifespan retrieval curve typically takes a form similar to that shown in Figure 14.1 (did your own lifespan retrieval curve take this form?). As can be seen in Figure 14.1 the lifespan retrieval curve consists of three components: the period of childhood amnesia (from birth to approximately five years of age), the period of the reminiscence bump (from 10 to 30 years) and the period of recency (which declines from the present back to the period of the reminiscence bump).



**Figure 14.1** An idealized representation of the lifespan retrieval curve

## 2.1 Childhood amnesia

There are many theoretical explanations of the period of childhood amnesia (see Pillemer and White, 1989), but most flounder on the fact that children below the age of five years have a wide range of specific and detailed autobiographical memories (Fivush *et al.*, 1996). Explanations that postulate childhood amnesia to be related to general developmental changes in intellect, language, emotion, etc., fail simply because apparently normal autobiographical memories were in fact accessible when the individual was in the period of childhood amnesia. It seems unlikely that an

increase in general functioning would make unavailable already accessible memories.

From the perspective of the Conway and Pleydell-Pearce (2000) model of autobiographical memory this period is seen as reflecting changes in the working self's goal hierarchy. The goals of the infant and young child, through which experience is encoded into memory, are so different, so disjunct, from those of the adult that the adult working self is unable to access those memories. Another possibility, one much more in line with Freudian thinking on childhood amnesia (Freud, 1955, first published in 1899), is that the working self of infancy and early childhood is much less able to control the occurrence and intensity of emotional experience. Episodic memories encoded during this period are then saturated with intense emotions and, if recalled in maturity, could destabilize the adult working self by reinstating intense infant emotions. This view suggests that access to autobiographical memories encoded during this period might be quite powerfully limited by the adult working self, leading to the lack of memories from this period. Currently, however, there is no generally accepted explanation for this component of the lifespan retrieval curve. Although it is not as mysterious as it once was, the period of childhood amnesia continues to present a challenge to autobiographical memory researchers.

## 2.2 The reminiscence bump

The second, and also very interesting, component of the lifespan retrieval curve, is the period when rememberers were aged 10 to 30 years, which is known as the **reminiscence bump** (RB). The RB is distinguished by an increase in recall of memories relative to the period that precedes it and those that follow it. (Was it present in your curve?) The RB has been observed in dozens of studies leading David Rubin (a leading researcher in this area) to describe it as one of the most 'reliable' empirical observations in cognitive psychology (Conway and Rubin, 1993; Rubin, 2002). Nonetheless, care must be taken in collecting memories for the RB. If memories are given dates as they are recalled then rememberers have a tendency to become 'stuck' in a time period. Then they may not produce a RB, or may produce an exaggerated RB depending on which time period the rememberer adheres to. Similarly, some rememberers can become 'stuck' in the very recent past and recall only memories from the last few months, again obscuring the RB and period of childhood amnesia.

In general, these types of retrieval strategies need to be minimized and access to memories should be open (rather than constrained or directed). The rememberer should therefore respond with the first memories to come to mind, i.e. those that are most accessible, for the full lifespan retrieval curve to be observed. When these conditions are met the RB is frequently observed. Interestingly, however, the RB is present not just in the recall of specific autobiographical memories but also emerges in a range of different types of autobiographical knowledge. For example, the RB has been observed in the recall of: films (Sehulster, 1996); music (cf. Rubin *et al.*, 1998); books (Larsen, 1998); and public events (Holmes and Conway, 1999; Schuman *et al.*, 1997). Memories recalled from the period of the RB are more accurate (Rubin *et al.*, 1998). They are judged as more important, by the individual, than memories from other time periods, and are rated as highly likely to be

included in one's autobiography (Fitzgerald, 1988, 1996; Fromholt and Larsen, 1991, 1992; Rubin and Schulkind, 1997). The autobiographical memories of middle-aged and older adults are therefore characterized by a high degree of accessibility to autobiographical memories dating to the period when they were 10 to 30 years of age, and this is typically most marked for the narrower period 15 to 25 years of age.

In a rather similar manner to the period of childhood amnesia the RB also has several plausible explanations (see Rubin *et al.*, 1998). Some obvious explanations can, however, be ruled out. Memories from the RB period are not dominated by first-time experiences, but rather appear to consist of memories of experiences that are idiosyncratic to individual rememberers. Similarly, the suggestion that memories from the RB are more vivid – the idea being that memory encoding is at peak efficiency during this period – turns out to be incorrect (Holmes and Conway, 1999; Rubin and Schulkind, 1997). Also incorrect are explanations that either postulate preferential effort to recalling memories from this period, or suggest RB memories are of more pleasant experiences. It has been found that no special effort is made to recall RB memories, and they are not of more pleasant events than memories from other parts of the lifespan (Rubin and Schulkind, 1997).

Instead, it seems that a more complex explanation is required and two candidate explanations currently exist. One comes from Rubin and colleagues (Rubin, 2002), a central hypothesis of which is that 'Events from the bump period are remembered best because they occur when rapid change is giving way to relative stability that lasts at least until retrieval' (Rubin, 2002, p.14). By this view a period of rapid change is dominated by novel experiences which more fully engage encoding processes, and so become represented in memory in a highly accessible way and lead to the RB. An alternative to the 'novelty' hypothesis is that the high accessibility of memories from this period may be related to their enduring relation to the self (Conway and Pleydell-Pearce, 2000). Possibly, many memories from the period of the RB are of **self-defining** experiences (Fitzgerald, 1988, Singer and Salovey, 1993), and have a powerful effect in binding the working self to a specific reality. The 'novelty' of RB experiences lies in their newness and uniqueness *for the self*, and they may play a crucial role in the final formation of a stable self system during late adolescence and early adulthood. Memories from this period help to define identity (Conway, 1996) and, because of this, they endure in memory in a highly accessible form. Which of these two hypotheses (the novelty hypothesis, or the self hypothesis) is the correct explanation is currently unknown but, as with so many supposedly 'alternate' explanations in psychology, it may turn out that both are required in order to develop a full theoretical account of the RB.

## 2.3 Recency

The final component of the lifespan retrieval curve, the 'recency' component (see Figure 14.1) can be simply explained as a period of forgetting older memories: memories recently encoded remain accessible, memories retained over a longer retention interval are subject to decay and/or interference and so become progressively less accessible. This pattern of retention is familiar from laboratory studies and is one that has been observed many times. On the other hand it might be questioned why such memories or salient experiences should be 'forgotten' in this

way. Moreover, it might also be noted that when people are specifically instructed to recall older autobiographical memories, there are apparently plenty of available memories (see Holmes and Conway, 1999, for example). Thus, what is of importance here is not the forgetting but rather a bias or preference in access. It may be that the recency portion of the lifespan retrieval curve reflects a lowering in self-relevance of memory for experiences from the recent past and, hence, a corresponding lowering of accessibility rather than complete forgetting. Thus, as recently acquired autobiographical memories become less relevant to the working self's goals, their accessibility is attenuated, but not lost, and can be restored by direct attempts to retrieve recent information. Of course, as the retention interval lengthens access may actually become lost, rather than just attenuated, and in that case forgetting would occur.

## Summary of Section 2

---

- The 'lifespan retrieval curve' illustrates how frequently autobiographical memories are recalled over different periods in someone's life. The lifespan retrieval curve is characterized by periods of childhood amnesia, the reminiscence bump and recency.
- The concept of the 'working self' (Conway and Pleydell-Pearce, 2000) can be thought of as a hierarchy of currently active goals and self-concepts through which experience is encoded and memories constructed.

## 3 Autobiographical knowledge, episodic memory, the working self and memory construction

### ACTIVITY 14.2

---

#### Taking part in an autobiographical memory experiment: retrieving memories to cue words

Before reading further, it will be useful to retrieve a few more of your own memories, and reflect on what comes to mind both while forming a memory and when it is fully constructed. To do this, imagine that you are a participant in an autobiographical memory experiment. The experimenter tells you that you will have to bring to mind memories of specific experiences of events that you yourself experienced, and that took place over periods of seconds, minutes, hours, but no longer than one day – as with the example memories listed at the start of the chapter. This means that responses such as 'last summer', 'when I was little', or 'holiday in Italy' are too general and do not count as memories. Instead you are required to recall detailed memories, memories of specific events. These can be

from any part of your life, indeed sampling widely would be good, but they should not be of events experienced in the last 12 months.

You are asked to recall specific memories by reading ‘cue’ words, then bringing to mind the first memory about which the cue word reminds you – bearing in mind the constraints of sampling widely and not from the past 12 months. Once you have the memory in mind, write down a description of it and provide a title. You should also rate each memory on the following scales:

**Table 14.1** Memory vividness

	1	2	3	4	5
Memory vividness	No imagery	Some imagery	Usual image vividness	Vivid imagery	Extremely vivid imagery
Valence of the remembered experience	Very negative	Negative	Neutral/Positive	Very positive	
Emotional intensity of the remembered experience	Very mild	Some emotion	Emotional	Intense	As intense as any emotional experience I have ever had
Rehearsal. How frequently have you thought and/or talked about this event?	Very rarely	Sometimes	With about average frequency	Above average	Very frequently

Okay let's create a response sheet now. On a sheet of paper write the following:

<b>Memory 1</b>					
<b>Title:</b>					
<b>Memory description:</b>					
<b>Ratings (circle a number):</b>					
Vividness:	1	2	3	4	5
Valence:	1	2	3	4	5
Intensity:	1	2	3	4	5
Rehearsal:	1	2	3	4	5

**AaE:** \_\_\_\_\_ (leave this blank for now)

Do this three times so you have three memory response sheets (in an actual experiment far more memories would be collected, usually 20 or more).

Assuming you are now ready:

- Recall your first memory to the cue word CHAIR and then complete the response sheet.
- Now recall a memory to the word ILLNESS and complete the response sheet.
- Finally recall a memory to the cue SUMMER and complete the response sheet.

Now, go back and at the bottom of each response sheet on the line that says 'Age' write (in months and as exactly as you can) your age when the remembered event occurred (Age at Encoding or 'Ae').

Keep the response sheets handy while you read the rest of this chapter, as we will often refer back to them. For now try to answer the following questions. Keep a record of your answers and come back to them when you have finished the chapter.

- (i) Did your memories always contain both abstract autobiographical knowledge as well as very detailed records of actual experiences?
- (ii) Were the details always or predominantly in the form of visual mental images? If not, what form were they in?
- (iii) Did you feel any emotions? Particularly with respect to the last two memories in comparison to the first memory.
- (iv) Did the memories just 'pop' into mind when you read the cue words, or did you have to elaborate the cue, for example think about a chair at home and some incident associated with it, such as when you bought it?
- (v) Did it take longer to retrieve a memory to cue two than to cues one and three?
- (vi) How complete a record of the actual event would you say the memory is?
- (vii) Did you notice how 'time compressed' the memory was? That is, it almost certainly took far less time to recall the event than the experience itself took.
- (viii) How accurate, as a record of the experiences, were the memories?
- (ix) At some point in the attempt to retrieve/construct the memories you must have decided that you had an appropriate memory in mind. Was this associated with any feelings? Did you have a sort of 'aha' experience when the memory came to mind? Did you feel as though you were almost reliving at least a small part of the past (memory researchers call this *recollective experience*)?

The pattern of memories retrieved over the lifespan has a particular shape, as shown in Figure 14.1, and one which strongly implicates the self in memory retrieval. The lifespan retrieval curve is, however, just one aspect of this complex higher order form of cognition. Another and equally important aspect is the *constructive* nature of autobiographical remembering. We know from the experience of our own memories that when knowledge of the past comes to mind, intentionally or spontaneously, it often features facts about ourselves and our lives, images of people, locations, activities and, of course, detailed (episodic) memories of specific events may be recalled (as in the cue word experiment you have just completed). It is this coming together of conceptual autobiographical knowledge, generic images and episodic memories that is the major form of construction in autobiographical remembering. In this section the nature and organization of autobiographical knowledge in long-term memory is considered first, followed by an account of episodic memories. The role of the working self in memory construction is then reviewed and, finally, the process of memory construction itself is outlined.

### 3.1 Autobiographical knowledge

One way in which autobiographical knowledge has been thought about is in terms of event specificity. Two broad types of autobiographical knowledge have been identified along this dimension: general events and lifetime periods (Conway and Pleydell-Pearce, 2000).

#### 3.1.1 General events

General events, as the term implies, are more strongly event-specific than lifetime periods, but not as event-specific as sensory–perceptual episodic memories, which are directly derived from actual experience (Conway, 2001) (see the discussion of episodic memories in Chapter 8). **General events** refer to a variety of autobiographical knowledge structures such as single events (e.g. the day we went to London), repeated events (e.g. work meetings), and extended events (e.g. a holiday in Spain). General events may themselves be organized in several different ways. For example, there may be ‘mini-histories’ structured around detailed and sometimes vivid episodic memories of goal-attainment in developing skills, knowledge and personal relationships (Robinson, 1992). Some general events may be of experiences of particular significance for the self and act as reference points for other associated general events (Pillemer, 1998; Singer and Salovey, 1993). Other general events may be grouped together because of their emotional similarity (McAdams *et al.*, 2001), and it is likely that there are yet other forms of organization at this level that await investigation (Brown and Schopflocher, 1998). However, the research currently available indicates that organization of autobiographical knowledge at the level of general events is extensive, and it appears to virtually always refer to progress in the attainment of highly self-relevant goals.

Conway and Pleydell-Pearce (2000), in a review, conclude that general events contain knowledge about locations, others, activities, feelings and goals common

to an event, as well as some specific episodic memories that help organize the general event knowledge. This autobiographical knowledge may be represented in several different ways and consist of images, feelings, verbal statements, associated together in a mental model (cf. Johnson-Laird, 1983 and Chapter 12). However, autobiographical knowledge in general events predominantly takes the form of generic visual images, i.e. images derived from repeated experiences (Brewer, 1986, 1988, 1996; Conway, 1996, 2001; Rubin and Greenberg, 1998). General event autobiographical knowledge can be used to access associated sensory-perceptual episodic memories and, when it is used in this way, a specific and detailed autobiographical memory can be formed. Thus, a specific AM will usually, if not always, contain some general event knowledge and this will often be in the form of generic images (was this the case with the memories you recalled earlier?).

### 3.1.2 Lifetime periods

In one of the few studies of this type of knowledge Robinson (1992) examined people's memories for the acquisition of skills (e.g. riding a bicycle or driving a car) and for aspects of personal relationships. These general events were found to be organized around a series of vivid memories relating to **goal attainment**. Consider two examples from Robinson's study: 'Ever agreeable, and eager to do anything that would get me out of the doldrums of inferiority, my father rented a bike and undertook to help me to learn. ... I shall always remember those first few glorious seconds when I realized I was riding on my own ...' (Quinn, 1990, cited in Robinson, 1992, p.224).

The first time I flew an airplane was one of the best firsts. It marked a sense of accomplishment for myself, and it also started me on the career path I have always wanted to follow. The day was warm and hazy, much as summer days in Louisville are. My nervousness didn't help the situation, as I perspired profusely. But as we took off from runway 6 the feeling of total euphoria took over, and I was no longer nervous or afraid. We cruised at 2,500 feet and I worked on some basic manoeuvres for approximately 45 minutes. We then returned to the airport, where I realized that this will soon be a career.

(Robinson, 1992, p.226)

These 'first time' memories can cue other related memories and the whole general event carries powerful self-defining evaluations that persist over long periods of time. Importantly, Robinson found many memories featured goal-related evaluative knowledge or self-defining memories (Singer and Salovey, 1993) along with more general knowledge and specific episodic memories. General event autobiographical knowledge can also be used to access related **lifetime periods** that contain associated knowledge. Lifetime periods, like general events, contain representations of locations, others, activities, feelings, and goals

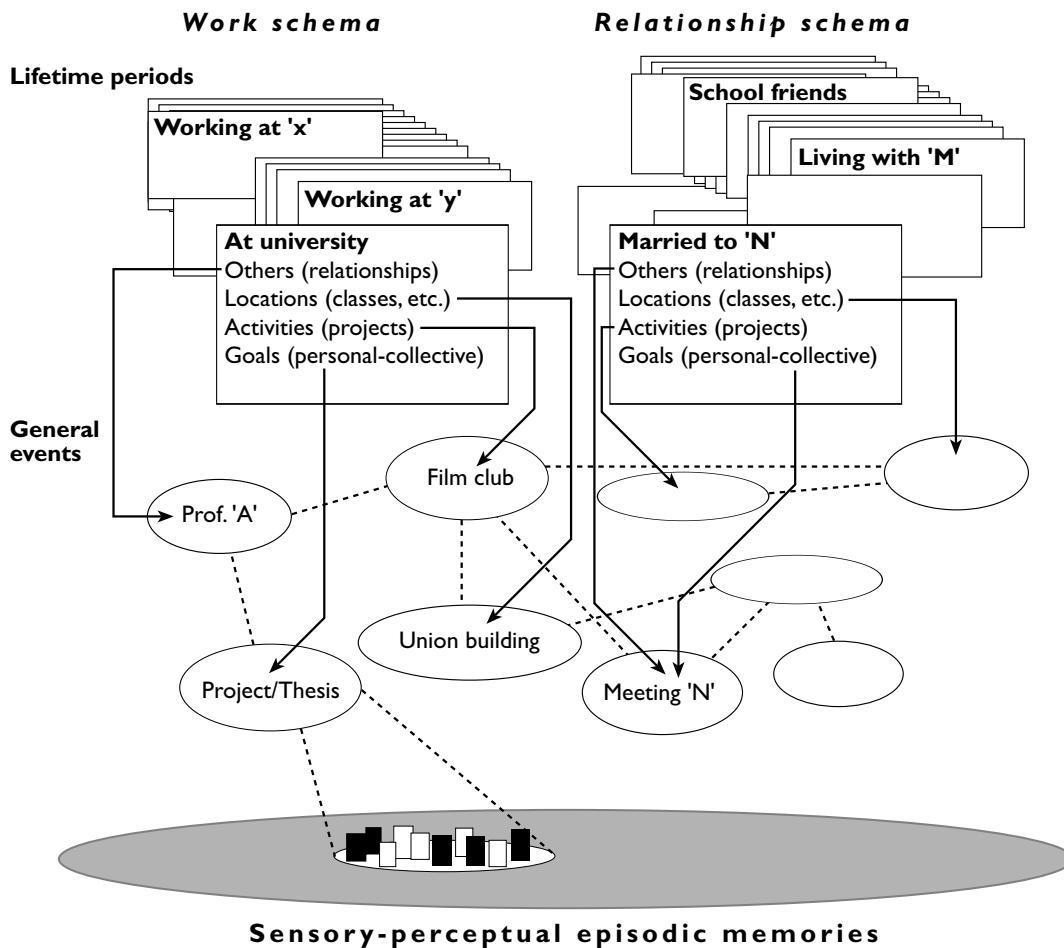
common to the period they represent. They effectively encapsulate a period in memory and in so doing may provide ways in which access to autobiographical knowledge can be limited, channelled or directed. As with general events there is evidence that lifetime periods contain evaluative knowledge, negative and positive, of progress in goal attainment (Beike and Landoll, 2000), and it seems likely that lifetime periods may play an important role in what Bluck and Habermas (2000) call the **life story**.

A life story is some more or less coherent theme or set of themes that characterize, identify and give meaning to a whole life. A life story consists of several life story **schema**, which associate together selective autobiographical knowledge to define a theme (Bluck and Habermas, 2000). A schema is a memory structure that encapsulates an event such that common parts are fixed, while variable parts occur as ‘slots’. Thus a schema for ‘going to the cinema’ would have pre-defined common parts (such as queuing for tickets, buying popcorn) and slots for variable parts (which cinema we went to, who I was with, what film we saw). Lifetime periods might provide the autobiographical knowledge that can be used to form life story schema and thus support the generation of themes. This may be particularly so because of the goal-evaluative information they contain. For example, a lifetime period such as ‘when I was at university’, will consist of representations of people, locations, activities, feelings and goals common to the period, but will also contain some general evaluation of the period, i.e. this was an anxious time for me, living away from home was difficult, I was lonely, I found the work too difficult, etc.

Lifetime period evaluations access related general events and, in turn, episodic memories that, when formed, provide the ‘evidence’ justifying the evaluations (see Beike and Landoll, 2000, and Conway and Pleydell-Pearce, 2000 for more on how autobiographical knowledge ‘grounds’ the self in memories of experience). They could also form the basis of a life story schema and, in the example above, ‘when I was at university’, this might perhaps centre on the unsuitability of the individual to higher education. This in turn might support a theme of an individual more suited to ‘practical’ as opposed to ‘academic’ activities (cf. McAdams, 2001). Thus, lifetime period autobiographical knowledge is less event-specific than general event autobiographical knowledge, it is also more conceptual and abstract. It encapsulates significant parts of the life story and may form an important bridge from autobiographical memory to core aspects of the self. Figure 14.2 (overleaf) depicts this scheme of autobiographical knowledge organization, and shows how such knowledge may be represented at different levels to form hierarchical **partonomic** knowledge structures. Partonomic refers to the way that a specific episodic memory is *part of* a general event, which in turn is *part of* a lifetime period, which is part of a life schema (Conway, 1996).

## Life story

### Life story schema



**Figure 14.2** Autobiographical knowledge and episodic memories

### 3.2 Episodic and semantic memory

In Tulving's (1972) original distinction between episodic and semantic memory, the defining feature of episodic memory was that it contained spatio-temporal information (see Chapter 8, Section 3.1). Episodic memories were of specific events that occurred at unique times, while semantic knowledge was of abstract, conceptual, context-free knowledge not linked to any specific place, time or learning episode. The reference for episodic memory was then individual personal experience, whereas the reference of semantic knowledge was social and universal (Tulving, 1983). So, for example, if you now remember Activity 14.1, in which you recalled some memories, that is an episodic memory of part of the experience of reading this chapter. In contrast, recalling that two broad classes of knowledge in

long-term memory are termed ‘episodic’ and ‘semantic’ memory, with the former referring to memory for experiences and the latter to memory for conceptual knowledge, is a form of semantic memory. Attractive though this fractionation of long-term knowledge into episodic and semantic memory may be, it unfortunately has problems.

One problem is that episodic memories must, of course, contain semantic knowledge and this raises the question of how easily the two types of knowledge can be distinguished. A second problem is that there are knowledge representations in long-term memory that on Tulving’s (1972) original definition we would want to term ‘semantic’, but which contain spatio-temporal information. For example, a schema for ‘breakfast’, which specifies the location, time, actions, order of actions and objects involved of a typical breakfast (Schank and Abelson, 1977): is this a semantic or episodic representation? A third problem relates to autobiographical knowledge. For instance, a person may know that last year they took a holiday in Italy – no other information needs to be brought to mind. But the reference of this knowledge (namely holiday) is both personal and universal and, moreover, it clearly contains spatio-temporal knowledge (see also Dritschel *et al.*, 1992). The fourth problem is that Tulving himself has revised the concept of ‘episodic’. In its latest incarnation the distinguishing feature is that episodic memories when recalled cause recollective experience, i.e. the feeling of experiencing the past and this does not occur when other types of long-term knowledge are brought to mind (Wheeler *et al.*, 1997).

The episodic–semantic distinction is then a difficult one to sustain and this is especially true when we consider autobiographical memory. However, in an attempt to retain the concept of episodic memory, Conway (2001) put forward a revised view of the concept that was closer in spirit to Tulving’s original conception, but which sought to refine it to meet the main points of later criticisms and revisions. According to this new view, knowledge contained in episodic memories is very largely *sensory–perceptual* in nature. Figure 14.2 conveys this by depicting episodic memories in an undifferentiated pool of representations of sensory–perceptual experiences. Thus, episodic memory is a repository of ‘experience-near’, highly event-specific sensory–perceptual details of recent experiences: experiences that lasted for comparatively short periods of time (seconds, minutes or at most hours). These sensory–perceptual episodic memories do not endure in memory unless they become linked to more permanent autobiographical memory knowledge structures. Conway (2001) argues that access to sensory–perceptual episodic memories is rapidly lost. This is because most episodic memories do not become linked to more stable and permanent autobiographical knowledge in long-term memory and, as a consequence, they rapidly decay and become permanently inaccessible. As a simple demonstration, cast your mind back over the events of today. They will be extremely detailed and numerous. If you try the same exercise, remembering today’s events, in a day or so, or perhaps next week, few episodic memories will have been retained relative to the number available on the day of experience, although rather more may be retained in the way of general event autobiographical knowledge. Only those episodic memories integrated with or consolidated in long-term memory close in time to the actual experience will later be accessible and available to enter into the subsequent formation of autobiographical memories.

### 3.2.1 Recollective experience

Experience-near sensory-perceptual knowledge when accessed during memory construction supports **recollective experience** and, consequently, episodic memory has a unique affinity for this type of memory awareness (Wheeler *et al.*, 1997). Recollective experience is the sense or experience of the self in the past and is induced by images, feelings and other memory details that come to mind during remembering (see Gardiner and Richardson-Klavehn, 2000, for a review). This memory awareness or feeling state (the sense of the self in the past) signals to a rememberer that the mental representation it is associated with is in fact a memory of an experience that actually occurred, and is not a fantasy, dream, plan or some other (experience-distant) mental construction, such as a general event. Thus, recollective experience effectively says ‘this mental representation is a memory of an event experienced by the self’. Note that it does not follow from this that recollective experience always indicates a true memory – ‘true’ that is in the sense that the recalled experience actually occurred – but when recollective experience is present the probability is high that the remembered event was one that had been previously experienced (Conway *et al.*, 1996; Roediger and McDermott, 1995).

## 3.3 The working self

Constructing an autobiographical memory is a complex form of cognition and has several effects on processing generally. One of the main effects is that the entire cognitive system enters what Tulving (1983) called **retrieval mode**. In retrieval mode attention, or part thereof, is directed inwards towards internal representations of knowledge, and conscious awareness becomes dominated by these representations. As a memory is formed the rememberer’s awareness becomes emotionally influenced by recollective experience and a powerful sense of the self in the past arises. The division of attention that then occurs gives rise to an attenuation of all other cognitive processes and, because of this, recall of AMs could, potentially, be highly dysfunctional in that current processing sequences would be disrupted. In extreme cases, such as in the involuntary and intrusive recall of prior trauma that is symptomatic of PTSD, autobiographical recall may be pathologically disruptive to everyday functioning (as in memory 5 at the beginning of this chapter). The point being that constructing a specific and detailed AM is a major cognitive occurrence with consequences for all other types of processing. Memory construction has therefore to be controlled and according to Conway and Pleydell-Pearce (2000) this is one of the main functions of the working self (see also Markus and Ruvolo, 1989).

### 3.3.1 Goals and the working self

The working self is conceived as a complex hierarchy of interconnected goals, all of which are in varying states of activation, but only some of which can enter consciousness (see an extended discussion of goals in relation to the ACT-R cognitive architecture, in Chapter 16). The working self may also contain representations of at least some goal-related knowledge, e.g. lifetime periods, life schema and life story or stories, as well as currently active models of the self. It is through the working self goal structure that episodic memories are formed and autobiographical knowledge is abstracted from experience. Thus, goal-related experience is prioritized in terms of encoding, consolidation, accessibility and

construction into specific, if transitory, autobiographical memories. Strong evidence exists showing that overall goal orientation of particular personality types acts to raise the accessibility of goal-related autobiographical knowledge and so facilitate their recall. This work has its origin in a seminal paper by Markus (1977) who found that people with a strong personality trait relating to the dependent-independent dimension showed preferential access to memories of experiences in which they had behaved in dependent or independent ways. In contrast, individuals within whom the dependent-independent dimension was weak did not have this memory bias.

These types of self-memory congruency effects have since been observed in several studies and most especially in the work of McAdams into power, intimacy and generativity (McAdams, 1982, 1985, 2001; McAdams, *et al.*, 1997). McAdams (1982), using the Thematic Apperception Test, TAT (Murray, 1938, 1943), in order to assess nonconscious aspects of personality (McClelland *et al.*, 1989), categorized individuals (on the basis of their TAT responses) into those with a strong intimacy motivation or, in contrast, with a distinctive power motivation. Content analysis of subsequently free recalled memories of 'peak' and other experiences found that the intimacy motivation group recalled peak experiences with a preponderance of intimacy themes compared to individuals who scored lower on this motivation, who in turn showed no memory bias. Similarly, the power motivation group recalled peak experiences with strong themes of power and satisfaction. Interestingly, neither group showed biases in memories for more mundane, less emotional, less self-defining memories. These striking biases in memory availability by dominant motive type suggest that the goal structure of the working self makes highly available those aspects of the knowledge base that relate most directly to currently active goals. In more recent work McAdams *et al.* (1997) have examined the influence of the Eriksonian notion of 'generativity' on the life stories of middle-aged adults (Erikson, 1950). Generativity refers to nurturing and caring for those things, products and people that have the potential to outlast the self. Those individuals who were judged high in generativity, i.e. who had a 'commitment' life story, were found to recall a preponderance of events highly related to aspects of generativity. In contrast, those participants who were not identified as holding a commitment story showed no such bias.

Work by Woike and her colleagues has further established the connection between personality and memory (Woike, 1995; Woike *et al.*, 1999). In the tradition of personality research deriving from Murray (1938) and McClelland (e.g. McClelland *et al.*, 1989), Woike identified implicit and explicit motives in a group of people who then recorded memorable events over a period of 60 days. According to McClelland *et al.* (1989), implicit motives are evident in preferences for certain types of affective experience such as 'doing well' for achievement and 'feeling close' for intimacy whereas explicit motives are present in social values and aspects of the self that can be introspected. A corollary of this view is that affective experiences should give rise to memories associated with implicit motives. Explicit motives, on the other hand, should lead to memories of less affective, routine experiences, more closely associated with self-description than with measures of implicit motives (i.e. TAT performance). This was exactly Woike's finding in both a diary study and in a laboratory-based autobiographical memory retrieval implicit/explicit motive priming experiment. In a subsequent study Woike *et al.* (1999)

investigated groups of individuals classified as ‘*agentic*’ (concerned with personal power, achievement and independence) or as ‘*communion*’ (concerned with relationships, interdependence and others). *Agentic* personality types are considered to structure knowledge in terms of ‘*differentiation*’ (the emphasis is on differences, separateness and independence) whereas *communal* individuals, in contrast, structure knowledge in terms of ‘*integration*’ (the emphasis is on similarity, congruity and interdependence). Across a series of studies, people with *agentic* self-focus were found to consistently recall emotional memories of events that involved issues of agency (mastery, humiliation) with their content structured in terms of differentiation. People with *communal* self-focus recalled emotion memories featuring others, often significant others, in acts of love and friendship, with the memory content structured in terms of integration. These findings clearly implicate the self (particularly the focus of the self) in determining recall and lend further weight to the suggestion that the working self influences access to sets of goal-related memories. (Reflecting on the content of your own memories how would you classify yourself – *agentic*? *communal*? Neither clearly one nor the other?)

In an intriguing study Pillemer *et al.* (1996; also Pillemer, 1998) investigated memory for specific educational episodes (memory 4, at the beginning of this chapter). The initial impetus for this work was the observation that autobiographies often contain accounts of highly specific events that were ‘turning points’ (**self-defining moments**) for the individual and that usually involved the adoption of a superordinate life goal that then determined much of the individual’s later activities. Pillemer *et al.* (1996) found that students and alumni were frequently able to report, in detail, highly vivid memories of interactions with professors and other teachers that had profoundly influenced their academic interests and, sometimes, the whole of their lives. These were often events in which superordinate long-term goals were adopted by the individual, e.g. to become a chemist, a writer, etc. Consider the following account by a postgraduate mature student of her first undergraduate Shakespeare class:

I was fascinated by the easy way the professor roamed through Shakespeare, by just the amount of knowledge he had. He seemed to know everything. In fact, after class, I asked him if he could identify a quote I had found about fencing, ‘Keep up your bright swords, for the dew will rust them.’ Immediately he said ‘Othello, Act 1 Scene 2, I believe.’ Which turned out to be exactly right. I wanted to know a body of literature that well. I’m still working on it.

(Pillemer *et al.*, 1996, p.330)

Of course, not all self-defining moments are positive and Pillemer *et al.* (1996, and Pillemer, 1998), in the only questionnaire study of these types of memories to date, list several other memories of more negative educational experiences that led to a subject being dropped, negatively conceived as ‘difficult’ or ‘boring’, and, in some cases, the emergence of negative conceptions of self as a poor or incompetent learner.

Singer and Salovey (1993) provide one of the main statements on the relation between goals and memories. A major finding in their study was that memories

associated with feelings of happiness and pride were strongly linked with goal attainment and the smooth running of personal plans (see also Sheldon and Elliot, 1999). In contrast, memories associated with feelings of sadness and anger were linked to the progressive failure to achieve goals. Singer and Salovey (1993) proposed that each individual had a set of **self-defining memories** that contained critical knowledge of progress on the attainment of long-term goals. Such goals, e.g. attaining independence, intimacy, mastery, and so on, may have been adopted as solutions to dominant self-discrepancies arising from childhood experiences (Strauman, 1996). Related to this, Thorne (1995) found that the content of memories freely recalled across the lifespan by 20-year-olds conformed to what she called ‘developmental truths’. Thus, memories from childhood very frequently referred to situations in which the child wanted help, approval and love, usually from the parents, whereas memories from late adolescence and early adulthood referred to events in which the rememberer wanted reciprocal love, was assertive, or helped another.

The notion of a ‘working self’ consisting of an active complex goal hierarchy is a useful way in which to understand the pattern of findings from the study of personality and autobiographical memory. The evidence points to a particular role for the working self, and that is to modulate access to knowledge in long-term memory and to control what new knowledge enters the knowledge base, i.e. which episodic memories are rehearsed and so become integrated with long-term knowledge structures. Note that none of this control need take place consciously, and the nature of the active working self’s control structures may make some long-term knowledge highly accessible, e.g. self-congruent knowledge, whereas other knowledge may be inhibited, e.g. self-incongruent knowledge. In terms of encoding, working self goal structures may nonconsciously direct attention and influence post-encoding processing, i.e. rehearsal, and in this way determine what is retained (Ross, 1989).

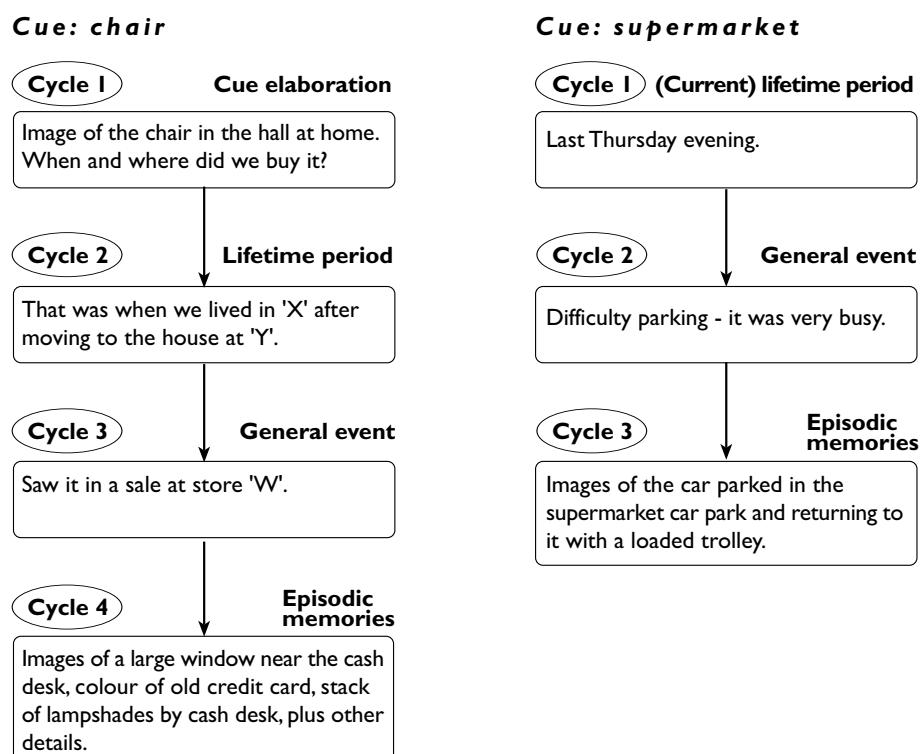
### 3.4 Constructing autobiographical memories

It has long been known that autobiographical memories can be intentionally constructed or, alternatively, may come to mind without the formation of any specific intention to recall a memory, i.e. to enter retrieval mode. We refer to the former type of construction as *generative retrieval* and the latter type as *direct retrieval*.

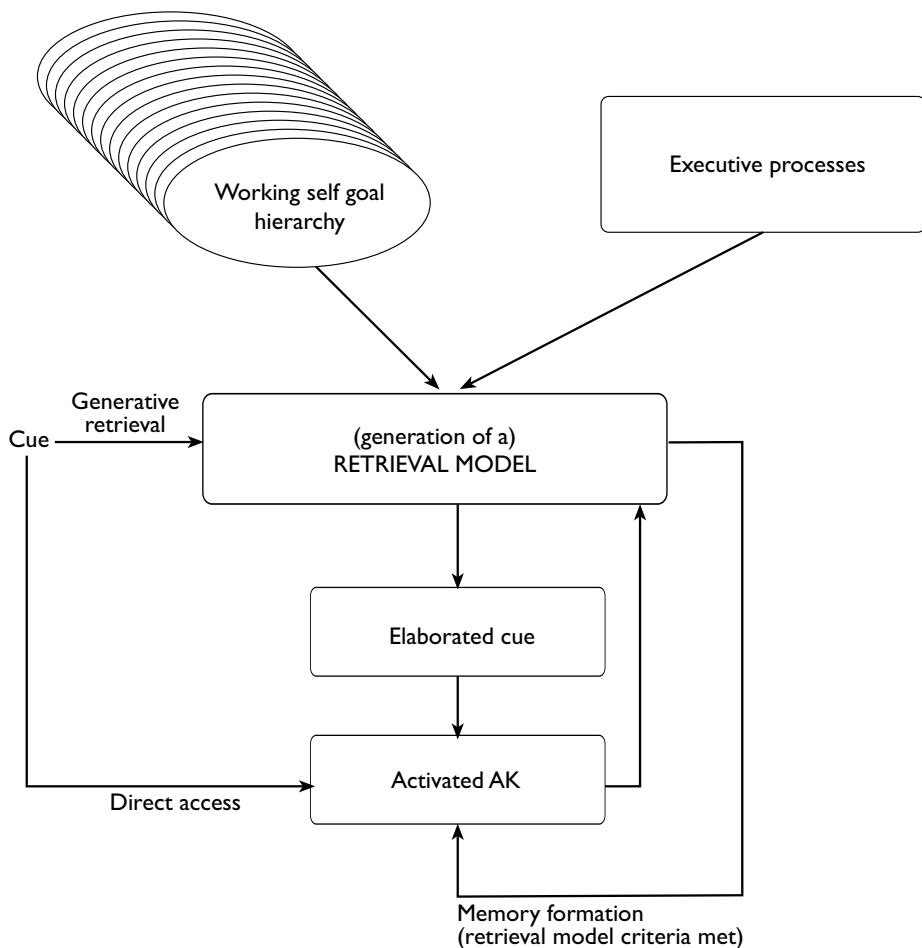
**Generative retrieval** occurs when remembering is intentional and the knowledge base is iteratively sampled as a memory is effortfully constructed. During this protracted process an initial cue is used to probe the knowledge base and accessed knowledge is evaluated against a retrieval model generated by the working self. If the constraints of the retrieval model are satisfied then a memory is formed, and the knowledge activated in the knowledge base (by the cue) together with associated goals of the working self form the autobiographical memory in that episode of remembering. Usually this process takes several or more cycles of access, evaluation and cue elaboration, as a stable pattern of activated knowledge that meets the constraints of the retrieval model gradually emerges. For example, in attempting to construct an AM to a cue such as ‘cinema’, a rememberer might elaborate the cue into the question ‘when did I go to the cinema a lot?’ This cue might lead to access of

the lifetime period ‘when I was a student’. Lifetime period knowledge can then be used to access general events, which in turn access episodic memories, and in this way a specific and task-relevant AM is constructed. Perhaps you were aware of this process when recalling memories to the cues ‘chair’, ‘illness’ and ‘summer’ in Activity 14.2? Figure 14.3 lists two protocols collected from people recalling memories to cue words while saying aloud what was going through their minds. Perhaps you were aware of similar types of knowledge coming to mind when you recalled your memories? Figure 14.4 provides a diagrammatic illustration of generative and direct retrieval.

Although the process of generative retrieval may seem laborious and is certainly effortful (retrieval times to word cues usually average between five and eight seconds), it nevertheless may operate with high efficiency when the system is in retrieval mode and multiple memories are to be recalled. Such circumstances would arise in a conversation with another person about a shared experience or in a discourse in which accounts of autobiographical memories form a part, e.g. in strategic self-disclosure, etc. Generally, however, recalling specific AMs is disruptive to other forms of cognition and, perhaps because of this, only occurs fluently under special conditions (intention to remember and retrieval mode). Indeed, the potential for disruption is great as autobiographical knowledge is highly cue sensitive, and patterns of activation across autobiographical knowledge structures in long-term memory continually arise and dissipate in response to



**Figure 14.3** Two protocols collected while rememberers related what came to mind when recalling memories to cue words



**Figure 14.4** Direct and generative retrieval

external and internal cues. These patterns dissipate over the components of general event and lifetime period knowledge structures (see Figure 14.2) but rarely spontaneously settle down into stable patterns that activate episodic memories. Nevertheless, when a cue activates a general event and associated episodic memories, a specific autobiographical memory can, apparently effortlessly and spontaneously, be formed: in other words **direct retrieval** occurs. In direct retrieval a cue causes a pattern of activation in autobiographical knowledge (AK) that stabilizes as a specific autobiographical memory and bypasses the stages of generative retrieval (repeated autobiographical knowledge access, evaluation and cue elaboration) (see Box 14.1 overleaf). Automatic awareness of the autobiographical memory does not necessarily follow and the working self can prevent (inhibit) a fully formed autobiographical memory from entering awareness (becoming linked to working self goal structures and initiating retrieval mode) if, for example, this would disrupt other processing that had a higher priority, i.e. for attaining a higher priority goal. One example of direct retrieval that enters awareness has been mentioned earlier, that is when patients with PTSD experience intrusive memories of a trauma that are involuntary triggered by cues linked to that event.

**14.1****Research study**

### **Haque and Conway's autobiographical memory 'probe' experiments**

An experiment by Haque and Conway (2001) illustrates how both types of retrieval occur when people recall specific autobiographical memories to a range of cue words naming common activities, locations and emotions. In this experiment the cue words were displayed on a computer screen and participants attempted to recall a memory to each cue individually.

In order to 'probe' the process of memory construction a signal was displayed two seconds, five seconds or 30 seconds after the cue word was on-screen. In response to the signal the participant had to report as exactly as they could the current contents of consciousness.

The reports were then classified for the predominant type of knowledge they contained, i.e. lifetime period, general event, specific memory or 'nothing in mind'. Table 14.2 (below) shows the number and percentage of each type of report at each of the probe intervals. From Table 14.2 it can be seen that similar numbers of reports at the two-second probe contain either autobiographical knowledge (lifetime periods and general events) or specific autobiographical memories.

Autobiographical knowledge indicates the operation of the generative retrieval process (a memory has not yet been formed) whereas the report of specific autobiographical memories at this very short probe indicates direct retrieval. Thus both types of retrieval can occur in the same individual. As can also be seen from Table 14.2, the incidence of reports of autobiographical knowledge at the longer probe times sharply decreases, although note the persistence of some general event knowledge, while the formation of specific memories strongly increases. As might be expected, as the retrieval time lengthens so the generative process runs its course and specific autobiographical memories were formed.

**Table 14.2** Distribution of protocols by protocol type and probe time in experiment I from Haque and Conway (2001)

<b>Protocol type</b>	<b>Time of probe</b>		
	Two seconds	Five seconds	Thirty seconds
Lifetime period	30(70%)	10(23%)	3(6%)
General event	18(32%)	23(41%)	15(27%)
Specific memory	33(19%)	59(35%)	79(46%)
Nothing in mind	19(63%)	8(27%)	3(10%)

The generation of autobiographical memories is complex. Which is, perhaps, not so surprising given the central nature of this type of memory to self. The basic idea is, however, relatively simple: a control process (the working self) modulates access to the autobiographical knowledge base (autobiographical knowledge and episodic memories). In the case of a cue that directly maps onto episodic memories, as part of a general event and a lifetime period, a stable pattern of activation is formed and a

specific autobiographical memory can then become linked to working self goals, at which point a memory enters conscious awareness. If the cue does not correspond directly to prestored knowledge then it needs to be elaborated and the autobiographical knowledge base will be iteratively sampled as outputs (activated knowledge) are evaluated and the cue elaborated, i.e. cue specificity is increased. In this way the generative retrieval process successively elaborates the cue and, in so doing, it channels activation into autobiographical memory knowledge structures until a stable pattern of activation is formed that satisfies working self constraints, i.e. that the memory should be about topic 'X' and should have features 'Y' for it to be accepted as a memory. Once this occurs a specific autobiographical memory is formed and can enter conscious awareness. This constructive process, despite being effortful and attention-demanding, works fluently in everyday cognition. It occurs outside conscious awareness, although some of the products of generative retrieval can be consciously experienced, e.g. cue elaboration and activated long-term knowledge.

## Summary of Section 3

---

- Autobiographical remembering of a specific episode is *constructive* in nature. It brings together autobiographical knowledge (general events and lifetime periods), generic images and episodic memories.
- Recently Conway (2001) suggests a re-conceptualization of 'episodic memory'. Accordingly, episodic memory is thought to consist of 'experience near', highly specific, sensory-perceptual details of recent experiences. Only those episodic memories that then go on to be linked to long-term memory will be available later to support the formation of subsequent autobiographical memories.
- One role of the working self is to control the process of autobiographical memory construction. This is because constructing an autobiographical memory is a major cognitive occurrence, and has consequences for all other types of processing. The working self can be thought of as a complex goal hierarchy, modulating access to knowledge in long-term memory, and controlling knowledge that can enter the knowledge base. Episodic memories are then interpreted in terms of the self.
- The process of memory construction itself can be either intentional (generative retrieval) or unintentional (direct retrieval).

## 4 Autobiographical memory in distress

As mentioned earlier (see memory 5, at the beginning of this chapter) probably the most outstanding form of direct retrieval occurs in the clinical disorder, PTSD, in which memories for traumatic experiences figure prominently (Brewin and Holmes, 2003, provide a recent review). In PTSD a range of symptoms are present but one

that is most marked is that of persistent intrusive thoughts and memories. Consider the case of John (see below).

### Case study

**NB** John is not the real name of this patient. Details of the case have been changed in order to protect anonymity.

‘John’ was seen for an emergency appointment with a psychiatrist due to his recent suicidal thoughts. The psychiatrist noted that three months previously he had seen a friend fall to his death from a building, but would not talk about it. He was then referred for assessment with a clinical psychologist. At the first meeting he appeared distracted, jumpy and low in mood. He had stopped work three months previously and spent all day at home. He did not like to leave his house, although felt safer going out in the dark. He could not listen to music. He reported feeling very tired as he frequently had nightmares. He also described being overwhelmed by mental images of his friend’s death, which he tried hard to push away from his mind. He had periods where he felt unreal and cut off from other people.

Although autobiographical remembering often involves an effortful and constructive process, in an individual like John it seems almost impossible to prevent memories of the trauma coming spontaneously to mind. It is as though these were directly retrieved despite the clear disruption they caused. Such direct retrieval of scenes from highly negative experiences is perhaps not so uncommon. Think back to a traumatic experience that you may have had, such as a car crash. Did it ever haunt you afterwards, with vivid images of the experience just ‘popping’ into mind? Or have you ever been to see a horror film and then found the next day that images of the worst scenes intrude into your mind? People with PTSD like John, and like the man who provided memory 5 (at the beginning of the chapter) during therapy for PTSD relating to his road traffic accident, may have numerous intrusive memories of a trauma in a single day. Often these cause destabilizing emotions like intense anxiety, guilt, fear, and often all of these occur when a trauma image intrudes uncontrollably into consciousness. Memory intrusions in PTSD are highly disruptive to other cognitive processes, they hijack attention and ongoing experience, and in so doing make even the most routine tasks difficult. The sufferer is constantly being thrown into ‘retrieval mode’ and this diverts attentional resources away from other goal-driven processes. It is perhaps one of the reasons that John withdrew from work and daily life: the goals of everyday cognition were just too difficult for him to attain while his attentional capacity was taken up with intrusive thoughts and memories.

PTSD is made up of several components (American Psychiatric Association, 1994): the traumatic event; response at the time of trauma; and subsequent psychological symptoms.

## 4.1 Traumatic event

First of all, the patient needs to report having experienced a ‘traumatic event’. This is typically a situation in which the individual experienced or witnessed actual or threatened death, or serious injury to self or others. Examples include natural disasters, sexual assault, road traffic accidents, physical attack and torture. Trauma in this context does not include the everyday use of the word trauma, such as having a ‘traumatic day’ at work. In John’s case, the trauma was seeing his friend fall down the centre shaft of a stairwell in a block of flats, to his death.

## 4.2 Response at the time of trauma

It is not only the experience of a trauma that contributes to a diagnosis of PTSD. The person’s reaction to the trauma is also critical. For example, soldiers fighting in an army or doctors performing surgery are exposed to death or serious injury routinely. While many people in such occupations do at some stage become traumatized, many do not. In comparison to the number of people who have experienced ‘a trauma’, community studies have estimated variable rates of PTSD occurring within the general population ranging from 1 per cent to 14 per cent (American Psychiatric Association, 1994). Clearly, not everyone who has a trauma goes on to develop PTSD. Thus, the second component of PTSD is that the person’s response to the trauma involved intense fear, hopelessness or horror (American Psychiatric Association, 1994). For example, if there were two people in a car that crashed into the side of a bus, narrowly missing a head-on collision, one person might think ‘I’m going to die’ and be intensely afraid. The other person might feel only mild fear and conclude ‘Phew! I’m so lucky this isn’t worse’. Only the first person would display a symptom of PTSD and, most probably, only this person would go on to develop the full range of PTSD symptoms. In other words the traumatic event has to be experienced as stressful in a major way and possibly the neurobiological stress response of a major release of glucocorticoids must also occur if PTSD is to follow.

While the American Psychiatric Association’s definition focuses on reactions of intense fear, hopelessness and horror, other intense emotions are also frequently experienced at the time of trauma such as anger or shame (Grey *et al.*, 2001). Conway and Pleydell-Pearce (2000) suggest that the perception of extreme danger and/or imminent death poses a fundamental challenge to the goal system of the working self. This is because the prospect of imminent death or extensive injury (physical or psychological) falls outside the range of plans, goals and self-images that constitute the working self. Thus, the trauma experience cannot be easily processed by the working self as it threatens the entire goal system, and because of this the experience cannot readily be integrated with autobiographical memory knowledge structures in long-term memory. On the other hand if the working self were to survive such trauma it would be highly useful from a survival perspective to retain a detailed record of what occurred. Thus, the tension is between either not encoding the (life) threatening experience (traumatic amnesia) or keeping a good record of it just in case one survives (a vivid, or ‘flashbulb’, memory). In fact, what is frequently observed in PTSD is, initially at least, a fragmentary often jumbled memory containing highly vivid details that are unordered with respect to their original order of experience, and often interspersed with extensive ‘gaps’ (islands of amnesia). Just the sort of compromise we might expect when the working self is

caught in the double bind of encoding and not encoding. Conway and Pleydell-Pearce (2000) suggest that in some cases the working self responds by encoding aspects of the traumatic experience in terms of all the then active working self goals (rather than by integrating the new knowledge into the autobiographical knowledge base). The net effect of this is that (episodic) memories of some selective moments of the trauma memories appear to be ‘burned’ into memory (see Brewin *et al.*, 1996, and Ehlers and Clark, 2000, for other models of PTSD).

In John’s trauma, he was on the stairs several floors below his friend when he slipped, so could not see him initially. John’s first reaction was intense fear that the banging sound meant impending danger. Then he experienced intense horror as he saw his friend fall past. He therefore meets the ‘emotional response’ criteria for PTSD. At the time John also felt intensely unreal, as if he was watching the event happen to him from outside of his body. He felt as if time had slowed down and that he was watching the event in slow motion. This phenomenon is known as detachment or dissociation (American Psychiatric Association, 1994), and is a common reaction during trauma. At the time of trauma, the feelings of unreality John experienced can be thought of as protecting the working self from the destabilizing psychological impact of the trauma (van der Kolk *et al.*, 1996). That is, through mentally distancing himself from the situation he may have protected the working self from being overwhelmed with emotion. Interestingly, it is also worth noting that while dissociation may be protective at the time of trauma it is also a strong predictor of developing PTSD (Shalev *et al.*, 1996).

### 4.3 Subsequent psychological symptoms

PTSD includes various psychological symptoms displayed by the patient. The symptoms present in three groups:

#### 4.3.1 Re-experiencing symptoms including intrusive memories

‘Re-experiencing’ a trauma includes having recurrent and intrusive recollections of the event (known collectively as ‘intrusions’), recurrent distressing dreams, having ‘flashbacks’ that involve suddenly acting or feeling as if the event were happening again, as well as intense physiological reactivity and psychological distress to reminders of the event. Ehlers *et al.* (2004) provide many case study examples of intrusive flashback memories; one of their examples neatly illustrates just how powerful these traumatic episodic memories can be in hijacking the entire cognitive system:

A patient who thought that he was going to die during an assault and would never see his children again, was not able to access the fact that he actually survived and saw his children again when he remembered this particularly distressing point of the assault. And when the intrusion occurred he would again be overwhelmed with sadness.

(Ehlers *et al.*, 2004)

It should be noted that these sensory–perceptual–affective details are just the sort of autobiographical knowledge contained in recently formed episodic memories in the

model outlined earlier. Indeed, it is central to the PTSD illness that the re-experience is mediated by episodic memories of the traumatic event itself. And this stands in sharp contrast to other psychological disorders that may also feature intrusive imagery of an *imagined* traumatic event (as in some cases of psychosis, e.g. of being cut in two by a man wielding a large sword, Morrison *et al.*, 2002) or of a catastrophic future event (as in Obsessive Compulsive Disorder, e.g. violently attacking elderly parents with an axe, de Silva, 1986). Thus, John, for example, was plagued by countless episodic images of his trauma, and mentioning the incident in his assessment interview caused him to re-experience them. He had five specific images that intruded. These were:

- a banging noise
- seeing his friend fall past him
- the bottom of a helicopter ambulance
- the exterior of the block of flats
- himself swearing at another friend.

These episodic images contained sensory and emotional experience from the time of encoding that included mood states at the time of trauma (for John, fear and horror) and feelings of unreality.

#### 4.3.2 Avoidance symptoms

Various reminders of the trauma trigger the re-experiencing of traumatic memories. In the exercise you did earlier you saw how cue words facilitate retrieval of detailed episodic memories. However, the key point about both generative and direct retrieval is that at some stage during retrieval a cue has to be present that can access the *content* of the sought-for knowledge (Tulving and Thompson, 1973). In PTSD, there may however be many cues for the trauma knowledge (especially if this is represented in memory in terms of goals active at the time of experience) and these have a tendency to generalize to other stimuli with shared features. For example, for a person raped by a bearded man, all men with beards may trigger re-experiencing symptoms, i.e. memory intrusions. Even more generally for someone who had a road traffic accident with a red car, the colour red, even on postboxes or clothes may trigger intrusive memories of the crash. These generalizations are made nonconsciously and at first it may not be evident why an intrusive memory comes to mind. Ehlers *et al.* (2004) give the following example: ‘A rape victim noticed that she was feeling extremely anxious while talking to a female friend in a restaurant and subsequently realized that the feeling was probably triggered by the presence of a man on another table who bore some physical similarity with the rapist’ (Ehlers *et al.*, 2004).

PTSD sufferers rapidly learn what triggers their re-experiencing intrusive memories and once learned, such potent cues are avoided, which can sometimes lead to dysfunctional behaviour, e.g. avoiding all red objects of a certain size, leaving a restaurant abruptly. **Avoidance** is then the second cluster of PTSD symptoms. John, for example, stopped playing all music as any rhythmic beat caused him to re-experience the ‘banging’ image and distress. He stopped going out during the day as the sight of tall buildings also brought back powerful intrusions. He also avoided

talking about the trauma, which made it difficult in therapy initially. Although avoidance may feel helpful in the short term, in the longer term avoidance of reminders of a trauma will not enable someone with PTSD to recover. Avoidance forms part of the vicious cycle which maintains the disorder (Ehlers and Clark, 2000).

#### 4.3.3 Amnesia as avoidance

A further form of avoidance is involuntary in nature and takes the form of amnesia for the trauma, or parts of the trauma. This may occur because the working self inhibits knowledge that was nonetheless encoded: often in therapy some memory returns to PTSD patients suggesting that it was in fact inhibited rather than not encoded. On the other hand the overwhelming of the working self by negative emotions may render encoding through this structure ineffective. The result is an amnesia more like that seen in anterograde amnesia following brain damage. This could also occur because the stress response causes a temporary increase in glucocorticoids and while levels of this neurohormone are raised the MTL (Medial Temporal Lobe) is temporarily disabled. At his therapy assessment John was not able to recall his memory of the time between seeing his friend fall past him at the top of the stairwell, and then when he was standing outside the building in a crowd. He said it felt like a ‘gap’ and, unlike the rest of his trauma, he was unable to recall it even with effort. Some patients can report having no awareness for hours after a trauma, and arriving in a place miles away with no idea of how they have got there – a dysfunctional and distressing consequence of **psychogenic** or **functional** amnesia (so termed to distinguish it from **organic amnesia** but, of course, all amnesia presumably has a physiological correlate).

#### 4.3.4 Hyperarousal symptoms

Hyperarousal may feel like being in a constant state of ‘red alert’ for potential danger. People with PTSD have an exaggerated startle response in that even small, unexpected noises make them jump. John repeatedly flinched throughout his assessment interview, for example if he heard a sound in the corridor outside. Other symptoms of hyperarousal are impaired concentration and irritability. Some patients are no longer able to concentrate on simple activities such as reading a newspaper or cooking. Impaired concentration links with features of autobiographical memory discussed earlier, that is, the process of retrieval of specific events (such as intrusions) is disruptive to other cognitive processes. John found it difficult to read the psychological assessment questionnaires and even to sustain attention when watching television.

### 4.4 Impact of symptoms

The final component of PTSD is the duration and impact of the disorder. To meet the diagnostic criteria the person must have had the cluster of PTSD symptoms for at least one month. This is because after a trauma most people typically have trauma symptoms, such as intrusive memories, for a short time. The impact criterion is standard to most mental disorders; that is, that the symptoms have caused significant and persistent distress for the person, for example in their occupation or socially. John had stopped working and also avoided all his friends. Indeed he was so

distressed by his symptoms and the impact they had had on his life, that he had begun to contemplate committing suicide.

In summary, then, to receive a diagnosis of PTSD, a patient must have experienced a traumatic event and responded to it with fear, helplessness or horror. The patient must persistently re-experience the trauma, avoid stimuli associated with it and have symptoms of increased arousal such as an exaggerated startle response. The trauma in John's case of PTSD was witnessing the violent death of a friend. John showed the full range of PTSD symptoms, especially intense intrusive memories, and these had a profound negative impact on his quality of life. Now we have gone through the diagnosis, go back and read the case vignette at the start of this section. Can you see how the symptoms John presented at assessment make sense?

#### **4.5 The nature of intrusive trauma memories**

Re-experiencing trauma in the form of intrusive memories is the hallmark symptom of PTSD, and because of it PTSD is one of the major psychopathological disorders of autobiographical memory. Phenomenologically, intrusive trauma memories have several distinctive features that relate to our understanding of normal non-traumatized autobiographical remembering. For instance, intrusive memories are image-based and they very often take the form of visual sensory-perceptual snapshots or 'film clips' (Ehlers and Steil, 1995). Just the sort of mental representation characterized earlier as sensory-perceptual-affective experience-near 'episodic' memories. Notably this contrasts with mental experiences associated with other anxiety disorders such as ruminative thoughts and worries, which often present in verbal form. The intrusive trauma images, although typically visual, may also incorporate sounds and smells and, sometimes, bodily sensations (Ehlers *et al.*, 2002). For example, a woman who was raped in the dark, encoded memory in non-visual sensory modalities, and during treatment reported suddenly experiencing physical pain and smells. John's images were a mixture of visual and auditory images. The memories also contain the emotion experienced at the time of trauma, such as fear, shame or disgust (Grey *et al.*, 2002) or feelings of unreality. As his friend fell past him, John felt unreal and saw the scene as if he were outside of his own body looking down on himself. He therefore experienced most of his intrusive images as if he was 'out-of-body' and often felt unreal. Also, intrusive memories can include verbal cognition from the time of trauma, typically catastrophic thoughts such as 'I'm going to die' (Holmes *et al.*, 2004).

Trauma memories have a quality of 'nowness' or 'live feel' (cf. Brown and Kulik, 1977). In normal autobiographical remembering the recall of specific events is accompanied by visual images and recollective experience, but in PTSD intrusive memories can be so compelling that the trauma feels as if it is in reality happening again. That is to say that the recollective experience component, the sense of the self in the past, appears to be overwhelmed or blocked by the intensity of re-experience: the PTSD sufferer does not have a sense or feeling of the self in the past, instead they are *actually* in that past moment. One woman with PTSD, for example, had a traumatic experience involving gun shots. Whenever she heard a sudden bang, such as a balloon pop at a children's birthday party, she would feel as if it was happening again and throw herself to the floor in protection. This type of experience is an example of a 'flashback'. Such cue-driven direct recall of

overwhelming re-experienced memories is unusual in normal autobiographical remembering. Nevertheless, the same mechanisms may be operating: namely a cue accesses the content of a sensory–perceptual–affective episodic memory and this becomes rapidly available to attention and consciousness. Without the working self to effectively intervene in direct retrieval the memory will capture attention and dominate consciousness. Perhaps, this would occur much more frequently in normal recall if control processes did not act to prevent patterns of activation, that constantly arise and dissipate in the knowledge base, coming to mind. Additionally, if episodic memories were more integrated with the autobiographical knowledge base it seems likely that they would not come to mind with such a feeling of ‘nowness’ and instead be recollectively experienced as a part of an extended and integral past. Indeed one of the goals of successful treatment of PTSD is to reach a point at which, when a patient recalls a trauma memory, it is experienced more like a normal autobiographical memory, as a part of the past and less as a part of the ‘now’ (Ehlers and Clark, 2000). In other words, the aim is to restore recollective experience.

Another point of departure between normal and trauma memories is that, unlike usual autobiographical remembering, the intrusive memories of trauma in people with PTSD often seem to be exact copies of what was experienced at the time of trauma. Moreover, the intrusions are usually highly consistent, being the same each time they come to mind. Such impressive consistency suggests that the same mental representation (episodic) memory is accessed each time an intrusion occurs. However, the veridicality of trauma images is a contentious issue and these may not always be based on experience. Holmes *et al.* (2004) found that, of a sample of patients with PTSD, approximately 2 per cent of different intrusive images were reported by participants as not actually being of their trauma experience. While this indicates that participants believed that most of their intrusions were of the event, it is possible that objectively they may not have been. Images, for instance, can be associated with beliefs that do not accurately reflect what in reality happened. Hackman *et al.* (2004) report the case of a woman who, after a house fire, experienced repeated intrusions of curtains burning. These led her to believe that her daughter was burning alive. However, she also had another intrusion of when she saw her daughter’s body in the morgue, which was unburned. The daughter had in fact died of smoke inhalation. The patient was, however, unable to connect the different information in the two images. Possibly these contradictory, highly vivid trauma images reflect unresolved affective conflicts in the patient.

Incidents of distortion and false images in trauma memories do then occur (Ehlers *et al.*, 2004, Conway *et al.*, 2004) and they pose an interesting question, namely: if a memory of a trauma is created in part to preserve a detailed record of what occurred when one survived a trauma, how can it contain distortions and errors? The question has not yet been addressed by the appropriate research but we might speculatively consider the role of the working self in generating phantasies that protect the self from deeply undermining cognition. Whatever eventually turns out to be the explanation it is clear that distortion in PTSD images require just as much an explanation as do accurate flashbacks.

Trauma images are like ‘highlights’ picked out in a trailer for a film, except that they can occur one by one rather than as a sequence and often do not appear to coalesce in any obvious way and instead present as fragmented and disorganized

(Foa *et al.*, 1995). Earlier it was reported that John had five distinct images of his trauma and these too appeared unorganized and fragmented. Interestingly, most PTSD patients spontaneously report between three to five trauma images (Hackmann *et al.*, 2004) suggesting that there may be some consistency in accessing trauma memories. More generally, however, one of the fascinating questions about trauma memories is why are there intrusive images of some moments but not others? Researchers have begun to address this question by investigating ‘hotspots’ in trauma memories (Richards and Lovell, 1999; Ehlers and Clark, 2000). These are elicited by asking the patient ‘what are the worst parts of your trauma when you describe it?’ These worst moments correspond to the moments that intrude. Grey *et al.* (2001) found that hotspot images are associated with a wide range of ‘peak’ emotions. They consist of the sensory–perceptual information encoded at that point in time, as well as the cognition linked to the specific emotion. For example, a patient had an extremely fear-filled image of the sound of impact during a crash and the sensation of being flung forwards, accompanied by the cognition ‘I’m going to die’. Hotspots that return as intrusive memories may then relate to moments during the experience of trauma when the working self was most intensely challenged and, clearly, repetitively and intrusively recalling such moments must act to destabilize the self.

## Summary of Section 4

---

- Direct retrieval in an extreme, disruptive and distressing form is illustrated by the intrusive memories of trauma (e.g. flashbacks) experienced by people with post-traumatic stress disorder (PTSD).
- The features of PTSD include the traumatic episode itself, the person’s experience at encoding, symptoms of re-experiencing the trauma in memory, avoidance and amnesia, and hyperarousal. The case study considered the clinical features from an autobiographical memory perspective.

## 5 Conclusion: what are autobiographical memories for?

A distinction is often drawn between ‘correspondence’ and ‘coherence’ models of memory. Correspondence models take as fundamental the accuracy of memory and its capacity, i.e. how much can be accurately remembered. Coherence models, in contrast, are not greatly concerned with accuracy and, instead, view the coherence of knowledge as being the fundamental principle guiding retention and remembering. The model of autobiographical memory described in this chapter draws on both these concepts and has as a central tenet what might be called adaptive coherence. Adaptive coherence always entails some degree of correspondence. Thus, episodic memories are summary records of short time-slices of experience and, to the extent that the experience accurately represented reality, then episodic memories are accurate records of the world. However, what is retained is filtered through the goal

structure of the working self and integrated with pre-existing, long-term memory knowledge structures and is, accordingly, highly selective. Furthermore, what is retained is not simply stored in some sort of passive way, like books on a library shelf, but rather is contextualized by autobiographical knowledge and brought into association with other autobiographical knowledge, at varying levels of specificity, when constructed as a memory. Episodic memories are then *interpreted* in terms of the self. And this brings us to the closing question of this chapter: what are autobiographical memories for? As we have seen, autobiographical knowledge and constructed memories serve many functions, as must be the case for such a central form of cognition, although ultimately autobiographical memory can be characterized as having one overriding function – it links, indeed it binds, the self to reality.

## Further reading

For reviews of autobiographical memory research see:

- Conway, M.A. and Pleydell-Pearce, C.W. (2000) ‘The construction of autobiographical memories in the self memory system’, *Psychological Review*, vol.107, no.2, pp.261–88.
- McAdams, D.P. (2001) ‘The psychology of life stories’, *Review of General Psychology*, vol.5, no.2, pp.100–22.

For reviews of the major theories of PTSD and treatment see:

- Brewin, C.R. and Holmes, E.A. (2003) ‘Psychological theories of posttraumatic stress disorder’, *Clinical Psychology Review*, vol.23, no.3, pp.339–76.
- Foa, E.B., Keane, T.M. and Friedman, M.J. (eds) (2000) *Effective Treatments for PTSD*, New York, Guilford Press.

## References

- American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4th edn), Washington, DC, American Psychiatric Association.
- Baddeley, A.D. (1986) *Working Memory*, Oxford, Clarendon Press.
- Baddeley, A.D. (2000) ‘The episodic buffer: a new component of working memory?’ *Trends in Cognitive Sciences*, vol.4, no.11, pp.417–23.
- Beike, D.R. and Landoll, S.L. (2000) ‘Striving for a consistent life story: Cognitive reactions to autobiographical memories’, *Social Cognition*, vol.18, no.3, pp.292–318.
- Bluck, S. and Habermas, T. (2000) ‘The life story schema’, *Motivation and Emotion*, vol.24, no.2, pp.121–47.
- Brewer, W.F. (1986) ‘What is autobiographical memory?’ in Rubin, D.C. (ed.) *Autobiographical Memory*, Cambridge, Cambridge University Press.
- Brewer, W.F. (1988) ‘Memory for randomly sampled autobiographical events’ in Neisser, U. and Winograd, E. (eds) *Remembering Reconsidered: Ecological and*

- Traditional Approaches to the Study of Memory*, New York, Cambridge University Press.
- Brewer, W.F. (1996) 'What is recollective memory?' in Rubin, D.C. (ed.) *Remembering Our Past. Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.
- Brewin, C.R., Dalgleish, T. and Joseph, S. (1996) 'A dual representation theory of posttraumatic stress disorder', *Psychological Review*, vol.103, no.4, pp.670–86.
- Brewin, C.R., and Holmes, E.A. (2003) 'Psychological theories of posttraumatic stress disorder', *Clinical Psychology Review*, vol.23, no.3, pp.339–76.
- Brown, N.R. and Schopflocher, D. (1998) 'Event cueing, event clusters, and the temporal distribution of autobiographical memories', *Applied Cognitive Psychology*, vol.12, no.4, pp.305–19.
- Brown, R. and Kulik, J. (1977) 'Flashbulb memories', *Cognition*, vol.5, no.1, pp.73–99.
- Conway, M.A. (1996) 'Autobiographical memories and autobiographical knowledge' in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.
- Conway, M.A. (2001) 'Sensory perceptual episodic memory and its context: autobiographical memory', *Philosophical Transactions of the Royal Society – Series B – Biological Sciences*, vol.356, no.1413, pp.1375–84.
- Conway, M.A., Collins, A.F., Gathercole, S.E. and Anderson, S.J. (1996) 'Recollections of true and false autobiographical memories', *Journal of Experimental Psychology: General*, vol.125, no.1, pp.69–95.
- Conway, M.A., Meares, K. and Standart, S. (2004) 'Distortions of imagery in memories of traumatic experiences', *Memory* (in press).
- Conway, M.A. and Pleydell-Pearce, C.W. (2000) 'The construction of autobiographical memories in the self memory system', *Psychological Review*, vol.107, no.2, pp.261–88.
- Conway, M.A. and Rubin, D.C. (1993) 'The structure of autobiographical memory' in Collins, A.E., Gathercole, S.E., Conway, M.A. and Morris, P.E.M. (eds) *Theories of Memory*, Hove, Sussex, Lawrence Erlbaum Associates.
- de Silva, P. (1986) 'Obsessional compulsive imagery', *Behaviour Research and Therapy*, vol.24, no.3, pp.333–50.
- Dritschel, B., Williams, J.M.G., Baddeley, A.D., Nimmo-Smith, I. (1992) 'Autobiographical fluency: A method for the study of personal memory', *Memory and Cognition*, vol.20, pp.133–40.
- Ehlers, A. and Clark, D.M. (2000) 'A cognitive model of posttraumatic stress disorder', *Behaviour Research and Therapy*, vol.38, no.4, pp.319–45.
- Ehlers, A., Hackmann, A. and Michael, T. (2004) 'Intrusive re-experiencing in post-traumatic stress disorder; phenomenology, theory and therapy', *Memory* (in press).
- Ehlers, A., Hackmann, A., Steil, R., Clohessy, S., Wenninger, K. and Winter, H. (2002) 'The nature of intrusive memories after trauma: the warning signal hypothesis', *Behaviour Research and Therapy*, vol.40, no.9, pp.995–1002.

- Ehlers, A. and Steil, R. (1995) 'Maintenance of intrusive memories in posttraumatic stress disorder: a cognitive approach', *Behavioural and Cognitive Psychotherapy*, vol.23, no.3, pp.217–49.
- Erikson, E.H. (1950) *Childhood and Society*, New York, W.W. Norton and Company.
- Erikson, E.H. and Erikson, J.M. (1982/1997) *The Life Cycle Completed*, New York, W.W. Norton & Co.
- Fitzgerald, J.M. (1988) 'Vivid memories and the reminiscence phenomenon: the role of a self narrative', *Human Development*, vol.31, pp.261–73.
- Fitzgerald, J.M. (1996) 'Intersecting meanings of reminiscence in adult development and aging' in Rubin, D.C. (ed.) *Remembering our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.
- Fitzgerald, J.M. and Lawrence, R. (1984) 'Autobiographical memory across the life-span', *Journal of Gerontology*, vol.39, no.6, pp.692–8.
- Fivush, R., Hammond, C. and Reese, E. (1996) 'Remembering, recounting, and reminiscing: the development of autobiographical memory in social context' in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.
- Foa, E.B., Molnar, C. and Cashman, L. (1995) 'Change in rape narratives during exposure therapy for posttraumatic stress disorder', *Journal of Traumatic Stress*, vol.8, no.4, pp.675–90.
- Franklin, H.C. and Holding, D.H. (1977) 'Personal memories at different ages', *Quarterly Journal of Experimental Psychology*, vol.29, pp.527–32.
- Freud, S. (1955, first published 1899) 'Screen memories' in Strachey, J. (ed. and trans.) *The Standard Edition of the Complete Psychological Works of Sigmund Freud Volume 3*, London, Hogarth Press.
- Fromholt, P. and Larsen, S.F. (1991) 'Autobiographical memory in normal aging and primary degenerative dementia (dementia of the Alzheimer type)', *Journal of Gerontology: Psychological Sciences*, vol.46, pp.85–91.
- Fromholt, P. and Larsen, S.F. (1992) 'Autobiographical memory and life-history narratives in aging and dementia (Alzheimer type)' in Conway, M.A., Rubin, D. C., Spinnler, H. and Wagenaar, W. (eds) *Theoretical Perspectives on Autobiographical Memory*, Utrecht, Kluwer Academic Publishers.
- Gardiner, J.M. and Richardson-Klavehn, A. (2000) 'Remembering and knowing' in Tulving, E. and Craik, F.I.M. (eds) *Handbook of Memory*, Oxford, Oxford University Press.
- Grey, N., Holmes, E. and Brewin, C. (2001) 'It's not only fear: Peri-traumatic emotional "hot spots" in posttraumatic stress disorder', *Behavioural and Cognitive Psychotherapy*, vol.29, no.3, pp.367–72.
- Grey, N., Young, K. and Holmes, E. (2002) 'Cognitive restructuring within reliving: a treatment for peri-traumatic emotional "hot spots" in posttraumatic stress disorder', *Behavioural and Cognitive Psychotherapy*, vol.30, no.1, pp.37–56.
- Hackmann, A., Ehlers, A., Speckens, A. and Clark, D.M. (2004) 'Characteristics and content of intrusive memories in PTSD and their changes with treatment', *Journal of Traumatic Stress*, vol.17, no.3, pp.231–40.

- Haque, S., and Conway, M.A. (2001) ‘Probing the process of autobiographical memory retrieval’, *European Journal of Cognitive Psychology*, vol.13, no.3, pp.1–19.
- Hollway, W. and Jefferson, T. (2000) ‘Doing qualitative research differently’, *Free Association, Narrative and the Interview Method*, London, Sage.
- Holmes, A. and Conway, M.A. (1999) ‘Generation identity and the reminiscence bump: memories for public and private events’, *Journal of Adult Development*, vol.6, pp.21–34.
- Holmes, E.A., Grey, N. and Young, K.A.D. (2004) ‘Intrusive images and “hotspots” of trauma memories in posttraumatic stress disorder: emotions and cognitive themes’, *Journal of Behaviour Therapy and Experimental Psychiatry* (in press).
- Howe, M.L. and Courage, M.L. (1997) ‘The emergence and early development of autobiographical memory’, *Psychological Review*, vol.104, pp.499–523.
- Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge, MA, Harvard University Press.
- Larsen, S.F. (1998) ‘What is it like to remember? On phenomenal qualities of memory’ in Thompson, C.P., Herrmann, D.J., Bruce, D., Reed, J. D., Payne, D.G. and Toglia, M.P. (eds) *Autobiographical Memory: Theoretical and Applied Perspectives*, Mahwah, NJ, Erlbaum.
- Markus, H. (1977) ‘Self-schemata and processing information about the self’, *Journal of Personality and Social Psychology*, vol.35, no.2, pp.63–78.
- Markus, H. and Ruvolo, A. (1989) ‘Possible selves: personalized representations of goals’ in Pervin, L.A. (ed.) *Goal Concepts in Personality and Social Psychology*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- McAdams, D.P. (1982) ‘Experiences of intimacy and power: relationships between social motives and autobiographical memory’, *Journal of Personality and Social Psychology*, vol.42, no.2, pp.292–302.
- McAdams, D.P. (1985) *Power, Intimacy, and the Life Story: Personological Inquiries into Identity*, New York, Guilford Press.
- McAdams, D.P. (2001) ‘The psychology of life stories’, *Review of General Psychology*, vol.5, no.2, pp.100–22.
- McAdams, D.P., Diamond, A., de Audin, E. and Mansfield (1997) ‘Stories of commitment: the psychosocial construction of generative lives’, *Journal of Personality and Social Psychology*, vol.72, no.3, pp.678–94.
- McAdams, D.P., Reynolds, J., Lewis, M.L., Patten, A. and Bowman, P.T. (2001) ‘When bad things turn good and good things turn bad: sequences of redemption and contamination in life narrative, and their relation to psychosocial adaptation in midlife adults and in students’, *Personality and Social Psychology Bulletin*, vol.27, pp.472–83.
- McClelland, D.C., Koestner, R. and Weinberger, J. (1989) ‘How do self-attributed and implicit motives differ?’, *Psychological Review*, vol.96, no.4, pp.690–702.
- Morrison, A.P., Beck, A.T., Glentworth, D., Dunn, H., Reid, G., Larkin, W. and Williams, S. (2002) ‘Imagery and psychotic symptoms: A preliminary investigation’, *Behaviour Research and Therapy*, vol.40, pp.1063–72.

- Murray, H.A. (1938) *Explorations in Personality*, New York, Oxford University Press.
- Murray, H.A. (1943) *The Thematic Apperception Test: Manual*, Cambridge, MA, Harvard University Press.
- Pillemer, D.B. (1998) *Momentous Events, Vivid Memories*, Cambridge, MA, Harvard University Press.
- Pillemer, D.B., Picariello, M.L., Law, A.B. and Reichman, J.S. (1996) 'Memories of college: the importance of specific educational episodes' in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.
- Pillemer, D.B. and White, S.H. (1989) 'Childhood events recalled by children and adults' in Reese, H.W. (ed.) *Advances in Child Development and Behaviour Volume 21*, San Diego, CA, Academic Press.
- Richards, D. and Lovell, K. (1999) 'Behavioural and cognitive interventions in the treatment of PTSD' in Yule, W. (ed.) *Post-traumatic Stress Disorders: Concepts and Therapy*, Chichester, Wiley.
- Robinson, J.A. (1992) 'First experience memories: Contexts and function in personal histories' in Conway, M.A., Rubin, D.C., Spinnler, H. and Wagenaar, W. (eds) *Theoretical Perspectives on Autobiographical Memory*, Dordrecht, The Netherlands, Kluwer Academic Publishers.
- Roediger, H.L., III and McDermott, K.B. (1995) 'Creating false memories: remembering words not presented in lists', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.21, pp.803–14.
- Ross, M. (1989) 'Relation of implicit theories to the construction of personal histories', *Psychological Review*, vol.96, pp.341–57.
- Rubin, D.C. (2002) 'Autobiographical memory across the lifespan' in Graf, P. and Ohta, N. (eds) *Lifespan Development of Human Memory*, Cambridge, MA, MIT Press.
- Rubin, D.C. and Greenberg, D.L. (1998) 'Visual-memory-deficit amnesia: a distinct amnesic presentation and etiology', *Proceedings of the National Academy of Sciences*, vol.95, pp.1–4.
- Rubin, D.C., Rahhal, T.A. and Poon, L.W. (1998) 'Things learned in early adulthood are remembered best', *Memory and Cognition*, vol.26, pp.3–19.
- Rubin, D.C. and Schulkind, M.D. (1997) 'The distribution of autobiographical memories across the lifespan', *Memory and Cognition*, vol.25, pp.859–66.
- Rubin, D.C., Wetzler, S.E. and Nebes, R.D. (1986) 'Autobiographical memory across the adult lifespan' in Rubin, D.C. (ed.) *Autobiographical Memory*, New York, Cambridge University Press.
- Schank, R.C. and Abelson, R.P. (1977) *Scripts, Plans, Goals, and Understanding*, Hillsdale, N.J.: Erlbaum.
- Schuman, H., Belli, R.F. and Bischofing, K. (1997) 'The generational basis of historical knowledge' in Jodelet, D., Pennebaker, J. and Paez, D. (eds) *Political Events and Collective Memories*, London, Routledge.

- Sehulster, J.R. (1996) 'In my era: Evidence for the perception of a special period in the past', *Memory*, vol.4, pp.145–58.
- Shalev, A.T., Peri, T., Canetti, L. and Schreiber, S. (1996) 'Predictors of PTSD in injured trauma survivors: a prospective study', *American Journal of Psychiatry*, vol.153, no.2, pp.219–25.
- Sheldon, K.M. and Elliot, A.J. (1999) 'Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model', *Journal of Personality and Social Psychology*, vol.76, no.3, pp.482–97.
- Singer, J.A. and Salovey, P. (1993) *The Remembered Self*, New York, The Free Press.
- Strauman, T.J. (1996) 'Stability within the self: a longitudinal study of the structural implications of self-discrepancy theory', *Journal of Personality and Social Psychology*, vol.71, no.6, pp.1142–53.
- Thorne, A. (1995) 'Developmental truths in memories of childhood and adolescence', *Journal of Personality*, vol.63, no.2, pp.138–63.
- Tulving, E. (1972) 'Episodic and semantic memory' in Tulving, E. and Donaldson, W. (eds) *Organization of Memory*, New York, Academic Press.
- Tulving, E. (1983) *Elements of Episodic Memory*, Oxford, Clarendon Press.
- Tulving, E. and Thompson, D.M. (1973) 'Encoding specificity and retrieval process in episodic memory', *Psychological Review*, vol.80, pp.352–73.
- van der Kolk, B.A., van der Hart, O. and Marmar, C.S. (1996) 'Dissociation and information processing in posttraumatic stress disorder' in van der Kolk, B.A., McFarlane, A.C. and Weisaeth, L. (eds) *Traumatic Stress*, New York, Guilford Press.
- Wheeler, M.A., Stuss, D.T. and Tulving, E. (1997) 'Towards a theory of episodic memory: the frontal lobes and autonoetic consciousness', *Psychological Bulletin*, vol.121, pp.351–4.
- Woike, B. (1995) 'Most-memorable experiences: evidence for a link between implicit and explicit motives and social cognitive processes in everyday life', *Journal of Personality and Social Psychology*, vol.68, no.6, pp.1081–91.
- Woike, B., Gershkovich, I., Piorkowski, R. and Polo, M. (1999) 'The role of motives in the content and structure of autobiographical memory', *Journal of Personality and Social Psychology*, vol.76, no.4, pp.600–12.



Jackie Andrade

---

### 1 Introduction

Consciousness is probably the most fascinating and challenging subject of psychological research. Although we know that much of human cognition occurs at a subconscious level, most of us feel, rightly or wrongly, that it is our conscious thoughts that form our personalities and inspire our actions. Over centuries, philosophers have provided vocabularies for discussing the human mind, frameworks for investigating consciousness and possible solutions to some of the problems of consciousness. Only more recently have psychologists begun to research consciousness in its own right. Much of this recent research concerns the biological aspects of consciousness. It uses neuroscience techniques such as recording the electrical activity of the brain to discover how brain activity differs when we are conscious, or conscious of something, from when we are unconscious or unconscious of something. However, progress is also being made with cognitive approaches to consciousness. Cognitive psychology is helping to define the functions of consciousness, investigating how conscious processes differ from unconscious processes and suggesting possible evolutionary functions of consciousness. Consciousness research is still a frontier area of psychology. The different explorers still lack agreement about how to explain consciousness, or even how to define the problem, but they are making exciting discoveries.

This chapter aims to explain briefly the historical and philosophical roots of consciousness research, and then to discuss the place of consciousness as a concept in contemporary cognitive psychology. It then considers empirical studies of aspects of consciousness and cognitive accounts of consciousness.

It is difficult to give a coherent account of this topic because, although many areas of cognitive psychology inform our understanding of consciousness, these areas are not well integrated and have not been pulled together into a grand theory of consciousness. Much recent cognitive psychology research in the field of consciousness studies has focused on unconscious cognition (the terms unconscious, nonconscious, and implicit cognition are used interchangeably in much of the literature). Baars (1988) recommends contrasting conscious and unconscious cognition and using the differences between them to infer the functions of consciousness. He calls this procedure ‘contrastive analysis’. Many studies do not do this, however, but just focus on trying to demonstrate truly unconscious learning or memory. We shall look at some of these studies to see whether there is convincing evidence that we can remember or learn without being aware of doing so.

Many cognitive psychologists research high-level processes that are apparently dependent on consciousness – for example, visual attention, working memory, mental imagery. As an example of this research, we shall look in Section 2.2 at a relatively old study by Schneider and Shiffrin (1977) that helps show the conditions under which automatic and controlled processes operate. Automatic processes are relatively unconscious, in the sense that we have little awareness of their operation, whereas controlled processes are associated with conscious awareness of what is

being processed. The concept of controlled processing appears in the notion of working memory (see Chapter 9), where the central executive controls the operation of the phonological loop and visuo-spatial sketchpad, and in models of action selection, where controlled processing enables us to behave in novel ways rather than acting through habit. We shall discuss working memory and a model of action selection in Section 4 of this chapter when we consider different ways of explaining consciousness.

Dissociations between conscious and unconscious processes might suggest that we have a specific module, or modules, for consciousness. We shall consider how cognitive neuropsychology, the study of the effects of brain injury, can shed light on the issue of where if anywhere consciousness occurs. The main thrust of the chapter though is to explore the functions of consciousness. We shall look briefly at how studies of altered states of consciousness can complement more conventional studies of cognition in suggesting hypotheses about the functions of consciousness. I will argue that although consciousness appears to be associated with particular cognitive processes, for example selective attention, all we really know is that these processes are correlated with conscious awareness of stimuli in the environment or in memory. They are the **cognitive correlates of consciousness**. Discovering these correlates of consciousness does not explain conscious experience – why it feels the way it does to see blue or remember a face or imagine a voice – but it does help us to understand the possible role of consciousness in cognition. This role seems to include integrating selected information from different processing modules and making that information available across the cognitive system so that it can guide our behaviour.

The term ‘consciousness’ means different things to different people, so we begin by trying to define what it is we want to study.

## 1.1 Defining consciousness

### ACTIVITY 15.1

---

Before reading further, spend five minutes thinking about what it means to be conscious. Make a list of the special features of consciousness.

### COMMENT

Consciousness can be thought of in different ways. There is the state of consciousness, in the sense of being awake and aware of ourselves in our environment, rather than being asleep and more or less oblivious to what is happening around us. Also under this heading come altered states of consciousness brought about by drugs or hypnosis. Then there is consciousness in the sense of awareness of particular sensations or mental events. Thus, while reading this sentence, you may be conscious of someone entering the room or the taste of your coffee, but unconscious of the hardness of your chair or the hum of distant traffic. This sense of consciousness has been termed ‘access consciousness’. ‘Phenomenal consciousness’ refers to the particular qualities of our conscious experiences; what it feels like to taste coffee or hear the sound of footsteps, for example. Finally, there is self-consciousness, our awareness and monitoring of what we are doing, feeling, thinking, etc.

---

What was on your list of features of consciousness? For many people, one of the most salient aspects of being conscious is that we have a feeling of control over our behaviour and even over our thoughts. We feel we act in a particular way because we decided to act that way, that we have free will. Although it may be difficult to stick to our new year's resolutions, we can modify our behaviour in less ambitious ways. If you get a headache while reading this, you may decide to stop reading – and behave accordingly. Alternatively, you may choose to ignore your headache and attend to the chapter because you wish to finish reading it before going out. Resolving, deciding, choosing, ignoring, pain, attending, and wishing are also aspects of our conscious mental life.

Our conscious experience seems fairly continuous. William James (1918, first published in 1890) described it as a 'stream of consciousness'. We have a coherent and persistent awareness of ourselves and our environment, and are unaware of brief or inconsequential changes in our sensory input. For example, you were probably unaware of this page disappearing from view last time you blinked.

Consciousness is not only about our ability to control our behaviour or to know what is going on around us. It is also about feelings and experiences; for example, the smell of spices as you walk past a restaurant, the taste of chocolate, the sensation of jumping into a cold swimming pool or relaxing in a hot bath, the particular feel of looking at something red (rather than something green). Philosophers use the term 'qualia' to describe these qualitative, subjective, experiential aspects of consciousness.

Consciousness has so many different features that, before studying it, we need to know whether it is actually a single thing or several quite different things inappropriately called by the same name. Could a theory of consciousness in principle explain all the different aspects of consciousness that we have discussed so far, or will we need different solutions to different problems of consciousness? This chapter will say little about states of consciousness – being awake rather than asleep for example – although it briefly discusses what altered states of consciousness could reveal about the cognitive correlates of consciousness. It also says nothing about self-consciousness. Rather, it focuses on *consciousness of* particular stimuli or mental events; that is, awareness of particular sights, sounds, memories, ideas, mental images and so on.

Even if we limit our discussion to consciousness *of* things, there are still two aspects of this type of consciousness to consider. There is the consciousness itself, being aware rather than unaware of something, and there is the experience that this consciousness engenders, what it feels like to taste chocolate or perceive green for instance. Block (1995) argues that we should treat these two aspects of consciousness as separate problems. He uses the term **access consciousness** for the problem of how, when we are conscious of something, we are able to name it, remember it, decide whether to pick it up, etc., and **phenomenal consciousness** for the experiential aspects of the problem. The term 'access consciousness' captures the idea that the contents of consciousness are accessible to other cognitive processes; thus we can talk about our memories or remember things we said. Access consciousness describes this cross-talk between different cognitive modules (a module is a set of processes acting together and separately from other sets of

processes). Block argues that cognitive psychology only addresses the problem of access consciousness, despite sometimes claiming to solve the problem of phenomenal consciousness as well.

Chalmers (1996) makes a similar distinction. He refers to the problem of how information is shared between modular neural and cognitive systems as the '*easy problem*' of consciousness. Empirical research into vision, memory, attention, decision making and so on addresses the easy problem. The '*hard problem*' of consciousness, according to Chalmers, is to explain how and why the neural or cognitive processes of vision, memory, etc. give us the conscious experiences of seeing colours or enjoying happy recollections. In a similar vein, Levine (1983) argued that there is an '*explanatory gap*' between understanding the neural or cognitive basis of consciousness and explaining the phenomenology. There seems to be nothing about neural or cognitive processes that necessitates their being accompanied by particular experiences. Even if we knew everything about the structure and function of the visual system, could that ever be sufficient to explain why it feels the way it does to see red? This chapter discusses the extent to which cognitive psychology has helped advance our understanding of access consciousness.

## 1.2 Philosophical approaches to consciousness

This section gives a very brief introduction to philosophy of mind, so called because many of the issues pertain to mental processes and states in general, and not merely to conscious processes and states. This section is not intended to be a tutorial on philosophy of mind, just an overview of some of the philosophical issues facing researchers wanting to explain consciousness (in both the access consciousness and phenomenal consciousness senses).

Let's look at the problem of the explanatory gap more closely. Perhaps the reason it is so difficult to relate cognitive and neural brain processes to conscious experience is that they are two entirely different things. Here are three examples of the ways in which they appear to differ:

- 1 *Phenomenal quality.* Imagine looking at a particularly bright, warm shade of red. How can interactions between neurons or modules in your brain be 'bright' or 'warm' in the way your experience is? If you imagine a hot cup of black coffee, presumably nothing in your brain turns hot or black.
- 2 *Intentionality.* Philosophers describe conscious states such as desiring, believing, and perceiving as 'intentional', meaning they are about things. You can't just desire, you have to desire *something*. It is hard to see how brain states can be about things in the way that mental states are.
- 3 *Spatial position.* Neurons are physical entities so they take up space. One neuron can be to the left or right of another, but it does not make sense to talk about mental entities such as images or beliefs having spatial positions.

### 1.2.1 Mind/body dualism

Dualists such as Descartes solve this mind–body problem by arguing that the mind and the brain are entirely different things. The mind consists of an immaterial 'mindstuff' whereas the brain, like the rest of the body, is made of matter – water,

protein, lipids, etc. For consciousness researchers, this is a defeatist stance because it means that the mind does not obey natural scientific laws and is not amenable to scientific investigation. There are other strong objections to dualism. Perhaps the most important is that it does not explain how the mind interacts with the brain or body. How can a thought about drinking water make our physical hand move to pick up a glass and take a sip unless the thought is also somehow physical?

### 1.2.2 Monism

The converse of dualism is monism, the idea that mind and body are essentially the same thing. Philosophers and scientists who assume that consciousness is a property of the physical brain are called materialists. How do materialists deal with the explanatory gap? One way is to take an extreme view known as eliminative materialism. Proponents of this view argue that the apparent explanatory gap arises because we use mentalistic terms such as ‘desire’ and ‘belief’ which have no scientific basis. We should eliminate these terms from our scientific vocabulary and concentrate on investigating the underlying neuroscience of consciousness. They compare our current use of mentalistic terms with the use of the term ‘phlogiston’ (once thought to be a substance that escapes when matter burns), which was abandoned when new theories of natural science emphasized the role of oxygen in combustion. Most materialists do use mentalistic terms but argue that conscious states are brain states and concentrate on investigating their material basis – the chemical and neuronal interactions that underpin consciousness.

Functionalists, the vast majority of whom are also materialists, take a different approach to researching consciousness. Functionalism views mental states as functional or causal states, defined by the ways in which they transform some input (an external stimulus or the product of an earlier cognitive process) into output (information passed to another cognitive module or an overt behaviour). Conscious states are not just epiphenomena – mere by-products of brain processes that have no effect in themselves. Rather, they are the direct causes of our behaviour. Functionalists use the analogy of a computer: the brain is analogous to the hardware of the computer (the silicon chips, wires, etc.) and the mind is analogous to the computer’s software. The mind is implemented in (‘running in’, to use computer jargon) the physical brain, in the way that word-processing software might be implemented in a personal computer. The mind could also be implemented in some other physical system, just as a particular software package could run on different sorts of computers. A logical extension of this position leads us to ‘strong artificial intelligence’ (strong AI), the argument that, if we could program a computer with the same ‘software’ as a human, then it would be conscious in the same way as us. A less extreme position, weak AI, assumes that computers can have similar ‘mental’ properties to humans but that there might be something special about biological entities (e.g. carbon-based sensory systems) that make us conscious in the particular way that we are.

Functionalism lies at the heart of cognitive psychology. It means that cognitive psychologists can focus on investigating mental functions without too much reference to the brain biology that underpins them. Thus cognitive approaches to consciousness focus on explaining the mental processes that cause one conscious state or another, rather than investigating physical brain activity during that

conscious state. It is partly a question of finding an appropriate level of explanation for the phenomenon. Just as your success in an exam might best be explained in terms of your level of attention during lectures, the amount of rehearsal time devoted to your notes etc., rather than in terms of biological memory processes such as long-term potentiation, so might consciousness best be explained in terms of cognitive processes.

### 1.3 The place of consciousness within cognitive psychology

This section provides a brief reminder of the history of cognitive psychology, to help explain why cognitive psychologists are sometimes ambivalent towards the topic of consciousness, with some using it as a variable in their research but few studying it directly. Early in the history of experimental psychology, Wundt trained ‘observers’ to use introspection (to ‘look into’ their minds) to give detailed reports on their mental and emotional responses to stimuli. Introspectionism foundered partly because of the subjective nature of the data it produced. When two observers disagreed, it was not possible for an objective third person to resolve their disagreement by looking into their minds and deciding who was reporting their mental states more accurately, or indeed whether their mental states were the same or different. Other problems for introspectionism included the existence of unconscious, and hence unreportable, processes (e.g. the contribution of unconscious urges to adult behaviour that Freud stressed, and the unconscious processes in vision identified by Helmholtz), and a new emphasis on the functions rather than the structure of mental processes. For example, James (1918) suggested that the function of short-term memory was to keep in consciousness events that have just occurred.

These changes paved the way for a radical shift in the way human behaviour was studied. Behaviourists argued that psychologists should concern themselves with objective data, publicly observable behaviour, rather than subjective introspections. Although behaviourists could investigate what people said about their (mental) experience, speech being a form of behaviour, the emphasis was on studying the relationships between external stimuli and overt, behavioural responses. Consciousness itself was no longer a respectable topic for psychological research.

Cognitive psychology developed gradually from the middle of the twentieth century onwards, stimulated in part by a wartime need to explain the role of human factors in tasks such as radar monitoring and gunnery. It aimed explicitly to explain behaviour in terms of mental activity and thus represented a major shift in attitude towards the mind from behaviourism. Early cognitive theories included components that related to consciousness, such as attention, but they did not tackle the problem of consciousness directly and did not refer to conscious experience. One reason for this shyness of the topic may have been the need to be perceived as rigorously objective and scientific in an era still overshadowed by behaviourism.

Today, consciousness is increasingly considered as a variable in cognition, particularly in learning and memory research where conscious or explicit processing is contrasted with unconscious or implicit processing. Some examples are given in Chapter 8 on encoding and retrieval, as well as in Section 2 of this chapter. In the closely related field of neuropsychology, consciousness is also considered in

explanations of conditions such as blindsight. However, there is disagreement about whether the issue is really being tackled. Marcel (1988) argued that ‘reference to consciousness in psychological science is demanded, legitimate, and necessary’ (p.121). It is demanded because it is what makes our mental life interesting, what seems to make us who we are. According to Marcel, it is legitimate, because the concept of consciousness is no less coherent than other concepts in psychology such as intelligence or personality. It is necessary because we are often implicitly studying consciousness even if we profess to be more interested in some other aspect of cognition. For example, if we ask participants simply to press a button when a light flashes, we are still measuring their conscious experience. If they are not aware of the light flashing, they generally won’t respond. (This need not mean, however, that their conscious experience *caused* their response. It could be that the button push and the conscious experience are independent consequences of the nervous system’s processing of the light flash.) Despite Marcel’s call for more explicit discussion of consciousness in cognitive psychology, Banks (1993) argued that psychologists are still tiptoeing around the issue of consciousness rather as one might tiptoe around to avoid ‘waking the insane attic-bound Aunt of a Gothic novel’ (p.257). We might mention conscious processes such as attention, mental imagery or explicit memory, but we do not try to explain consciousness itself.

Despite the rise of behaviourism, introspection did not die out completely as a tool for psychological research. For example, ‘think aloud’ protocols have been used to study memory rehearsal and problem solving (see, for example, Section 1.2 in Chapter 10). However, introspection is becoming more widely used. Some of the studies discussed in the next section rely on participants’ reports of whether they were aware of experimental stimuli. Note that the current use of introspection usually only assumes that people have insight into the products of their cognitive processes, not that they can report the processes themselves.

## ACTIVITY 15.2

---

Think of cognitive theories from other chapters. What role does consciousness play in these theories? Do any of the theories help to explain consciousness?

### COMMENT

Although many cognitive theories include concepts like attention or working memory, they generally do not specify what processes or qualities make us conscious of some stimuli or cognitive products. For example, are we conscious of information by dint of it being in short-term memory, as James suggested? More recently researchers have argued that we are only fully aware of a subset of representations in working memory (e.g. McElree, 2001). The relationship between consciousness and working memory is discussed in Section 4 of this chapter.

---

## Summary of Section 1

---

- The term consciousness encompasses the state of being awake, our ability to control our behaviour and be aware of our surroundings, and our mental experiences or ‘qualia’.
- There is an explanatory gap between understanding the neural and cognitive functions of the brain and explaining conscious experience.
- Cognitive psychologists view mental states as causal states that affect our behaviour.

## 2 Empirical research: cognitive studies of consciousness

This section focuses on three areas of cognitive psychology. Each area tackles the problem of consciousness in part by investigating unconscious processes. Although this may seem perverse, it helps us work out what processes are associated only with consciousness and not with unconscious processing. Section 2.1 covers implicit cognition (specifically, implicit memory and learning, where there is no awareness of what is remembered or learned). Research into implicit cognition is important because it can help us to define consciousness better by contrasting it with unconscious processes. It raises the question of what, given the extent of unconscious processing, might be the function of consciousness. Section 2.2 revisits earlier research into automatic and controlled processing. Although not phrased in terms of unconscious and conscious cognition, these studies show us the essential characteristics of conscious processes. They are slow but flexible whereas automatic or unconscious processes are fast and efficient but inflexible. Section 2.3 considers briefly the neuropsychology of consciousness. Studies of conditions such as blindsight help elucidate the function of ‘normal’ consciousness and raise questions about the functional and physical structure of consciousness. This chapter necessarily misses much of the research in cognitive psychology that relates to consciousness. The areas it does cover are those in which researchers have particularly used their findings to frame questions about consciousness, although even in these areas much research is reported with scant if any mention of what it tells us about consciousness.

### 2.1 Implicit cognition

#### 2.1.1 Implicit memory

Implicit memory is memory without any accompanying sensation of remembering. It is revealed by changes in performance on specially designed memory tests. For example, if I show you the word ‘witness’ in the context of an apparently unrelated, non-memory, task and then give you a surprise memory test, you may not recall seeing ‘witness’ or recognize it as a word from the earlier task. However, if I ask you to say the first word that comes to mind starting ‘wit–’, your implicit memory for the

word will make you more likely to say ‘witness’ than if you had not just seen that word.

Tests such as this **word-stem completion task** are often referred to as indirect memory tests because they measure memory without directly asking people to decide if they remember the stimuli. Indirect tests are assumed to measure predominantly implicit memory, whereas direct tests measure explicit memory. Note though that no memory test is ‘process-pure’; performance on almost any memory test can be influenced by both implicit and explicit memory. For example, if you are asked to think of the first word that comes to mind beginning with ‘wit–’ and nothing comes to mind, you may try to think back to the earlier task to search for clues, for words that might fit the stem. If you remember that ‘witness’ was one of the words on the first task, and use that as your response, then you are using your explicit memory and the task is not giving a pure measure of your implicit memory.

Two studies of implicit memory are described below. Further examples are given in Chapter 8, but these two are chosen because they appear to show implicit memory in the absence of explicit memory (with the caveat about the process-impurity of memory tests).

In an early study of implicit memory, Eich (1984) showed that prior presentation of a word in a particular context could bias its subsequent interpretation. Participants in Eich’s experiment heard a list of word pairs. One word in each pair was a homophone – that is, it sounded like another word with a different meaning and spelling – for example, PANE (as in ‘pane of glass’) is a homophone of PAIN (as in ‘stomach pain’). The other word in each pair made clear the intended interpretation of the homophone. ‘Window-PANE’ and ‘taxi-FARE’ are examples of the word pairs used by Eich. Note that the homophones he used were the less common interpretations, PAIN and FAIR being the more frequently encountered spellings and meanings. Although participants in Eich’s study could hear the word pairs, they could not attend to them because their main task was to shadow (repeat) an essay played at the same time. Memory for the homophones was tested in two ways. On the recognition test, participants listened to a list of words and were asked to say whether each word was old (i.e. present in the unattended list of word pairs) or new. This is a direct test of memory because it requires participants to make a judgement about their memory; it is therefore assumed to measure mainly conscious or explicit memory. Participants were unable to recognize the unattended words. For the second memory test, the experimenter read out a list of words and participants were asked to spell them. Their spelling was biased by the previous presentation of the homophones (e.g. they were biased towards spelling P-A-N-E rather than the more common P-A-I-N). This spelling test is an indirect test of memory because it does not require deliberate recollection. These findings therefore suggest that participants had implicit memory for the homophones, but not explicit memory.

Another example of implicit memory in the absence of explicit recollection is the **false fame effect**. Jacoby *et al.* (1989) used a task on which participants had to say whether names belonged to famous people. In the study phase of Experiment 2 of their study, participants read aloud a list of 40 non-famous names. In the test phase, 10 of these names were mixed with 10 new non-famous names for a recognition memory test. The remaining 30 names were mixed with 30 new non-famous names and 60 famous names for the fame judgement task. Before the fame judgement task,

participants were told that the names they had just read were the names of non-famous people, hence if they recognized any name from the first phase of the experiment they should respond ‘non-famous’. The experimenters manipulated the degree of attention paid to the names in the study phase. In the full attention condition, participants were told that the experimenters were interested in their ability to pronounce the names quickly and accurately. In the divided attention condition, they were told to pay as little attention as possible to the pronunciation task, concentrating instead on listening to a stream of spoken digits and spotting runs of three odd numbers. The fame judgement task was sufficiently difficult that participants made a number of false positives, saying that a name was famous when in fact it was not. If participants had encountered a non-famous name earlier in the experiment, they were more likely to judge incorrectly that it was famous if the study phase took place under conditions of divided attention. Divided attention impaired explicit recognition of the names and thus reduced participants’ ability to use explicit memory to interpret feelings of familiarity. In the absence of conscious recognition, familiar names were assumed to be famous.

One of the most exciting things about this field of research was the discovery that people with amnesia often performed as well as people with normal memory on the indirect tests of memory. In other words, despite their severely impaired explicit memory, amnesics had almost normal implicit memory. For example, Squire and McKee (1992) replicated the false fame effect in amnesic participants. Amnesics were significantly impaired at recognizing the previously presented names, compared with the control subjects, but they were just as biased towards judging presented non-famous names as famous. Such findings led to new rehabilitation strategies because they showed that amnesics had the potential to learn new information even though they appeared to have no memory. We shall see an example of a new learning strategy that has been used in rehabilitation in Section 3 of this chapter (in Box 15.3 on errorless learning). The dissociation between the effects of brain injury on implicit and explicit memory raises questions about the structure of consciousness. We shall return to these questions in Section 2.3.

Implicit memory phenomena are often referred to as priming. Priming is the improvement in performance caused by previous exposure to the target stimulus or by previous or concurrent exposure to a closely related stimulus. For example, Meyer and Schvaneveldt (1971) showed that participants decided more quickly that pairs of letter strings were two real words when they were related words (e.g. ‘doctor’ and ‘nurse’) than unrelated words (e.g. ‘doctor’ and ‘cabbage’). Building on this study, Marcel (1983) showed that participants identified ‘doctor’ as a real word faster when it was preceded by a very brief presentation of the word ‘nurse’ than by an unrelated word. The word ‘nurse’ in this example is called the prime. The idea behind priming is that: (a) activation of an item’s representation in memory lingers, so that the representation is still slightly activated next time the item is encountered, making it easier to re-activate even if only a partial cue is presented, such as a word-stem; (b) activation spreads to representations of related items, making related representations easier to activate than unrelated, unprimed representations. Priming is also used in a more general sense, to refer to the activation of moods or stereotypes (see Box 15.1 on unconscious influences on behaviour). Demonstrations of priming show that we are not always aware of our

knowledge, or of the basis for our behaviour. Priming is often preserved even when brain injury causes impairments to explicit cognition; thus – as with other forms of implicit memory – people with amnesia typically have preserved priming. Box 15.3 on errorless learning (in Section 3) mentions a way of using this preserved priming in rehabilitation. Despite preserved implicit memory, lack of explicit memory has an impact on the more general aspect of consciousness that James (1918) termed our ‘stream of consciousness’. Baddeley (1990) cites the example of Clive Wearing, whose amnesia was so severe that he repeatedly noted in his diary that he had just regained consciousness. Memory and consciousness thus appear to be correlated, but the direction of causation is unclear: does normal consciousness require intact memory function or does normal memory function require consciousness? We shall return to this problem of correlational evidence in the conclusion to the chapter.

### 15.1

#### Unconscious influences on behaviour

Research in the field of social cognition suggests that priming may considerably influence our behaviour outside the laboratory. Primes may influence our mood and behaviour without us being aware of them. For example, Bargh *et al.* (1996) asked participants to arrange lists of words to form meaningful sentences. In the experimental group, each word list contained a word related to the concept of old age, for example ‘wrinkled’, ‘ancient’. Participants were surreptitiously timed as they left the laboratory after completing this task. Those in the experimental group, who had been exposed to the ‘elderly’ primes, left the laboratory more slowly than those in the control group, who were not exposed to those primes. Bargh *et al.* argued that the primes activated a stereotype of old age and participants behaved in accordance with that stereotype even though they had not noticed the primes.

Neumann and Strack (2000) showed that people’s mood can be affected by the mood of others around them, even when they are unaware of their mood change or its cause. When participants listened to text read in a sad voice, they were more likely to rate their own mood as sad but were unaware that their mood had changed as a result of listening to the sad voice.

Lieberman (2000) argues that implicit cognitive processes, such as priming of stereotypes and mood states, underlie the phenomenon commonly known as ‘intuition’; that is, our ability to judge social situations and respond appropriately without being aware of the information (other people’s moods etc.) on which we base our judgements.

#### 2.1.2 Implicit learning

Studies of implicit memory show that we can ‘remember’ things without having any conscious experience of remembering them. A more profound claim has been made by researchers in the field of **implicit learning**, namely that we can learn things without ever being aware of them. If this claim is true, it helps establish

some ground rules for our study of consciousness by telling us what is possible without consciousness. Evidence that a lot of learning is possible without consciousness would suggest that consciousness is just an epiphenomenon that plays no causal role in our cognition. We shall therefore look closely at some of the evidence for implicit learning and at some of the methodological problems that face researchers trying to show that participants had no awareness of the material they learned.

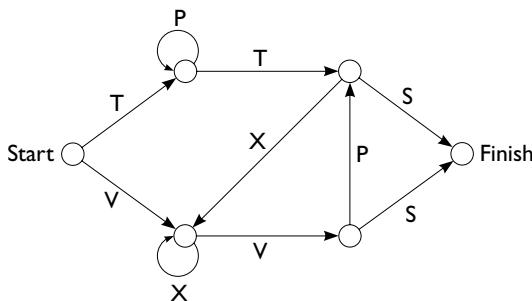
One way of demonstrating this unconscious or implicit learning is to present the stimuli to be learned very quickly, too quickly for participants to notice more than just a flash on the screen. This is called subliminal presentation. The study by Marcel (1983) mentioned in Section 2.1.1 has become a controversial classic: a classic because it demonstrated priming even though the presentation of the primes was apparently subliminal: and controversial because of claims that the findings could not be replicated when stricter definitions of ‘subliminal’ were used. However, subsequent researchers have used subliminal presentation with other test procedures to provide evidence of implicit learning.

Some of this evidence comes from demonstrations of the mere exposure effect, the tendency for people to prefer stimuli they have encountered before even if they were unaware of them during the previous encounter. For example, Kunst-Wilson and Zajonc (1990) presented novel black and white patterns very briefly. Even though participants said they could not see the patterns, because they were presented so briefly, they later tended to select those patterns when presented with pairs of patterns (one presented earlier and one new) and asked to choose the one they preferred. On a recognition test in which participants chose the pattern they remembered seeing earlier from each pair, participants performed at chance, i.e. they were just guessing. So their first, unconscious, encounter with the patterns apparently changed their emotional response to them even though they had no conscious or explicit memory for seeing the patterns before (see also Chapter 13 on cognition and emotion, Section 5.2.1).

As suggested by the controversy over Marcel’s (1983) study, there are problems with using subliminal presentation to demonstrate unconscious learning. One problem is ensuring that all stimuli are subliminal for all participants. This is tricky because some stimuli are easier to perceive than others. For example, you can sometimes hear someone say your name even if you don’t hear anything else they say because you are attending to another conversation (Moray, 1959). Another limitation is the equipment used to present the stimuli. Early studies used tachistoscopes, boxes that were specially designed to show stimuli for very brief and accurately timed periods. Many researchers now use computers for running their experiments, but computerized presentation times are limited by factors like the screen refresh rate – how quickly the computer redraws the display. A refresh rate of 17 ms means that stimuli can only be presented for multiples of 17 ms. This limitation makes it hard for the experimenter to present a stimulus for long enough to effect some learning but briefly enough to prevent the participant identifying the stimulus.

A solution to these problems is to present stimuli **supraliminally** (for long enough that they can be consciously perceived), but to test learning of some hidden relationship between them. For example, Reber has claimed that people can

implicitly learn hidden rules, constituting an **artificial grammar**, that underpin a set of supraliminally presented letter strings. Although they cannot verbalize their knowledge, it allows them to distinguish grammatical from ungrammatical items with above-chance accuracy. Two of Reber's early experiments are discussed below. An example of an artificial grammar is shown in Figure 15.1.



**Figure 15.1** The artificial grammar used by Reber (1967). Grammatical letter strings are created by following the arrows through the array from left to right. The curved arrows indicate items that can be repeated

### ACTIVITY 15.3

First, trace through the network shown in Figure 15.1 to convince yourself that TPPPTS and VXVPS are grammatical strings whereas VTPS and TPTTS are ungrammatical. Now try writing down or explaining to a friend the rules of this grammar.

### COMMENT

Verbal report is often used as a way of finding out whether participants in implicit learning experiments were aware of the information they learned. The stimuli used in these experiments are typically complex and novel and therefore difficult to describe. You may have struggled to report the rules of Reber's grammar even though you were looking at Figure 15.1 while doing so. The problem of eliciting participants' explicit knowledge through verbal report is addressed later in this section.

Reber (1967) asked participants to learn 28 letter strings or 'sentences' as part of a 'memory experiment'. The sentences were presented in seven sets of four sentences. Each sentence was viewed for 5 seconds, after which the participant tried to write it. After attempting to reproduce the four sentences in a set, the participant was told which sentences they had reproduced correctly and the procedure was repeated until they reached the criterion of reproducing all four sentences correctly on two consecutive trials. The next set of sentences was then presented. For participants in the experimental group, the sentences were formed according to the rules of an artificial grammar. For those in the control group, the sentences contained randomly ordered letters. The results were as follows. Both groups learned the second of the seven sets of sentences in fewer attempts than the first. However, from the third set onwards, the control group continued to make a mean of between eight and 11 errors

per set whereas the experimental group continued to improve, making a mean of only three errors on the seventh set. Reber argued that the experimental group acquired knowledge of the grammar that enabled them to learn more efficiently.

A second experiment in Reber's study showed that prior exposure to the grammatical sentences enabled participants to distinguish new grammatical sentences from ungrammatical sentences. Participants learned 20 grammatical sentences in a procedure similar to that described above. They were then tested on 88 trials with new sentences, comprising two presentations each of 22 grammatical sentences (that had not been encountered in the learning phase) and two presentations each of 22 ungrammatical sentences. Participants were told that the sentences they had already learned were grammatical, and were asked to use their knowledge of those sentences to decide if each test sentence was grammatical or ungrammatical. Their decisions were correct on a mean of 79 per cent of trials, well above the mean of 50 per cent expected from chance.

Subsequent studies of artificial grammar learning have used procedures similar to that of this second experiment. Participants learn a set of grammatical strings to a predetermined criterion, and then attempt to distinguish novel grammatical strings from ungrammatical strings. Performance is typically above chance, though not as impressive as in Reber's study, even though participants cannot state the rules they used to decide which were the grammatical items. This apparently implicit learning of the grammar is dissociable from explicit recall of the grammatical strings presented in the learning phase. For example, Knowlton *et al.* (1992) found that amnesic patients were as good as controls at classifying novel strings as grammatical or ungrammatical, but were poorer at recognizing exemplars that had been encountered in the learning phase.

Nissen and Bullemer (1987) used an alternative procedure for testing implicit learning of hidden regularities between visible stimuli. They gave control and amnesic participants a choice reaction time task in which they had to watch a panel of four lights (ABCD) and, whenever a light came on, to press the key under that light as quickly as possible. Participants were not told that the lights came on in a fixed order (the 10-item sequence DBCACBDCBA was presented repeatedly). Controls and amnesics got faster at this task until the sequence was switched to a random order: at this point, their reaction times increased. The amnesic participants showed no awareness that the lights had come on in a regular sequence. Control participants are also often unaware of the sequence, particularly if they perform the key-pressing task under conditions of divided attention.

A problem with these demonstrations of implicit learning is that we have no way of determining participants' awareness of the key stimuli or relationships while they are doing the task. If we ask them if they are aware of the grammar, for example, then we draw their attention to it and lose the opportunity for demonstrating learning without awareness. So researchers have to ask participants afterwards what they were aware of during the task. This is unsatisfactory, because it relies on people's memory of what they were aware of rather than measuring awareness online. Another problem is that quite a small amount of knowledge may be enough to boost people's performance above the chance level. For instance, they may not know the whole grammar, just that a certain letter can be repeated or come at the start of a letter string. Knowing possible starts for grammatical strings may be sufficient to

distinguish a few of the grammatical strings from the ungrammatical strings, resulting in performance that is slightly above baseline. If a test of awareness simply asks ‘were you aware of the grammar?’ or ‘what was the grammar?’, then it will miss the knowledge that actually boosted performance on the grammar test, and that knowledge may well be explicit. Participants may interpret the question as asking for a complete report of the grammar, which they cannot give, and so do not volunteer their knowledge of fragments of the grammar. These problems are discussed by Shanks and St John (1994). They argue that experimenters must use tests of explicit knowledge, or awareness, that meet two criteria before they can claim that learning resulted in truly implicit knowledge. The *information criterion* states that the test of awareness must probe for the sort of information that could support performance on the test of learning (for example, knowledge that a particular letter often comes at the start of a grammatical letter string). The *sensitivity criterion* states that the test of awareness must be sensitive to all the relevant explicit knowledge; it must be just as sensitive as the test of implicit knowledge. Simply asking participants to state the rules of the artificial grammar fails on both counts. It does not prompt them to report fragments of the grammar, thus failing the information criterion, and it does not give them any recall cues, thus failing the sensitivity criterion. The grammar judgement test is more sensitive because it presents the actual grammatical stimuli and these may serve as cues to memory.

Given the difficulty of ensuring lack of awareness of critical stimuli in awake participants, perhaps a better strategy would be to study learning in people who are unconscious. Testing patients receiving anaesthetics offers a way of tackling this issue, though one that presents more difficulties than might at first be imagined. One difficulty is that depth of anaesthesia, or ‘degree of unconsciousness’, fluctuates during an operation and there is not yet a universally agreed way of measuring this fluctuation or of establishing exactly the depth of anaesthesia at which a person loses consciousness, in the sense of losing all awareness of themselves and their surroundings. Thus a finding that patients can learn information presented during anaesthesia may reflect explicit learning during undetected moments of consciousness, rather than truly unconscious learning. Another difficulty is that the sensitivity of the memory tests has not been established. So, if a study shows no evidence for learning during anaesthesia, this may be due to use of a test that is too insensitive to detect small amounts of preserved learning. Not surprisingly, although many studies have investigated learning during anaesthesia, their findings have been mixed (Andrade, 1995).

Catherine Deeprose and I recently obtained evidence for priming during anaesthesia in a study that overcame some of the problems discussed above (Deeprose *et al.*, 2004). We played words (e.g. ‘tractor’) to patients during surgery. When the patients came round from the anaesthetic, we asked them to respond to word stems (e.g. ‘tra–’) with the first word that came to mind. Playing a word during surgery increased the likelihood of patients using that word to complete a word stem on recovery. In other words, they showed some implicit memory for the words even though they were anaesthetized while receiving them. We had pilot tested our word-stem completion test to ensure that it was reasonably sensitive and also reasonably uncontaminated by explicit memory (it was relatively unaffected by a manipulation of attention known to affect explicit memory). Thus we gave ourselves a good

chance of demonstrating implicit memory for words played during surgery. We minimized the chance of priming occurring during moments of awareness by using an EEG measure of depth of anaesthesia throughout word presentation and testing patients who were unparalysed, because the drugs that are often used to paralyse patients during surgery make it even harder to detect moments of consciousness. We are therefore reasonably confident that we have demonstrated that memories can be primed in someone who is unconscious. The next step is to investigate whether new information can be learned during anaesthesia.

To summarize, there is some evidence for learning without consciousness of what is learned. However, this implicit learning is rather difficult to demonstrate convincingly and is also not very useful, in the sense that we cannot revise, contemplate or tell people about what we have learned implicitly. We cannot select what we learn when learning implicitly, and we cannot retrieve the learned material voluntarily. It would appear that conscious processes (e.g. the active selection, rehearsal and elaboration of information) contribute to much of our everyday learning. Even so, implicit learning may help us to pick up repeated patterns or relationships among stimuli or events, and by doing so help us direct our conscious learning processes towards interesting features of our environment. Implicit memory is easier to demonstrate. When only a small amount of learning has occurred, because of inattention or brain damage for example, the resulting memory may be implicit. We may be unaware of what we have learned because the encoded material does not reach some threshold for consciousness or because it has not been processed by a ‘conscious memory module’. This issue of the structure of consciousness is discussed briefly in Section 2.3 on the neuropsychology of consciousness.

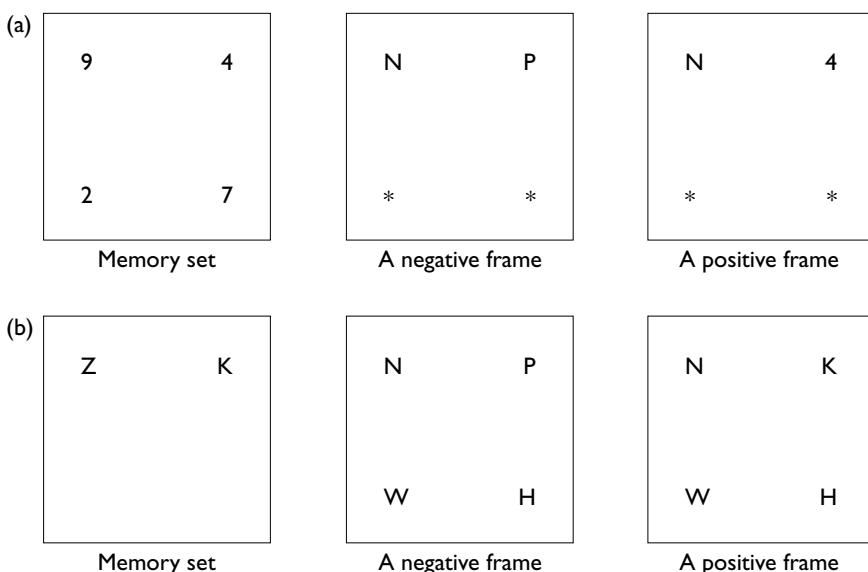
## 2.2 Controlled versus automatic processing

The concept of controlled processing is closely allied to that of conscious processing. As you have seen in the previous section, if we want to demonstrate implicit learning or memory, we have to make it very difficult for participants to process the target material in an active way, for example by distracting their attention from it. This sort of active processing is often known as **controlled processing** (as opposed to automatic processing). The idea of controlled processing is central to concepts such as working memory (discussed in Section 4 as a potential model of consciousness). Controlled cognitive processes typically accompany consciousness – that is, they are cognitive correlates of consciousness. This section therefore describes a classic study by Schneider and Shiffrin (1977) that defined and demonstrated controlled and automatic processes in visual attention.

Schneider and Shiffrin (1977) based their theorizing on Atkinson and Shiffrin’s (1968) model of memory (see Chapter 8, Section 2.1), arguing that automatic processes operate on the long-term memory store (an interconnected array of nodes) whereas controlled processes require the limited capacity short-term store, essentially the currently activated nodes of the long-term store. They defined automatic processes as the activation of a sequence of nodes in the long-term store via connections between those nodes that have become relatively permanent through repeated use. Once triggered, automatic processes operate without active control so it is difficult to stop them or change their course. In contrast, activation of a novel

sequence of nodes requires attention, which limits our capacity to activating just one novel sequence at a time but gives us control over the activation.

Schneider and Shiffrin demonstrated the difference between these two processing modes using ‘target search tasks’ that required participants to detect targets as quickly and accurately as possible from arrays of distractors. They manipulated (a) the number of targets participants had to search for; (b) the number of items (targets and distractors combined) on each slide or ‘frame’; and (c) the mapping between the set of targets for any series of trials (the ‘memory set’) and the distractor set. This mapping manipulation was the key to demonstrating automatic and controlled processing modes. In the **consistent mapping** condition, the targets were always selected from the same set of items and the distractors were always selected from a different set so, for example, participants might search for target digits among distractor letters. In the **varied mapping** condition, the targets and distractors were drawn from the same set, so participants might search for letters among letters and a particular letter could be a target on one trial and a distractor on another. Figure 15.2 gives examples of trials in these different conditions.



**Figure 15.2** Examples of stimuli from Schneider and Shiffrin’s (1977) experiments. On a given trial, participants saw and memorized the set of targets for that trial (the ‘memory set’). They then saw a sequence of frames. Their task was to detect whether any of the memorized targets appeared in the sequence: (a) shows stimuli from a consistent mapping condition with a memory set size of four and a frame size of two; (b) shows stimuli from a varied mapping condition with a memory set size of two and a frame size of four. The asterisks represent pattern masks in the positions not occupied by targets or distractors

The consistent and varied mapping conditions produced quite different patterns of response times. In the consistent mapping condition, performance was fast and relatively unaffected by memory set size and frame size. In the varied mapping condition, performance was slower and was slowed further still by increasing the number of targets to be searched for and the number of items per frame to be

searched. Schneider and Shiffrin argued that the varied mapping condition necessitated a controlled serial search through the array in each frame: hence, the larger the array, the longer the search. In contrast, the consistent mapping condition allowed the items in the arrays to be searched automatically, in parallel, because all we have to do is spot a digit (say) among letters, and we have had many years of practice at recognizing digits. Because we only have to spot a digit, there is no need to maintain the specific identities of the memory set items in short-term memory.

Subsequent researchers have extended Schneider and Shiffrin's description of controlled processes to cover a variety of cognitive processes that are flexible but slow and expensive in terms of cognitive resources. In everyday terms, we use controlled processing when performing novel tasks or wanting to override habitual behaviours. For example, when making black coffee for a friend, we may have to attend to each step of the procedure to stop ourselves adding milk. If talking to our friend at the same time, it is easy to lapse into automatic behaviour and make the 'action slip' of adding milk to their coffee as well as to ours. Automatic processes are typically fast and efficient but inflexible. We are generally unaware of the operation of automatic processes. For example, an experienced tennis player will hit a ball without stopping to think how to do so whereas a novice may have to think about how to hold the racquet, how hard to hit the ball and so on. Automatic processes make little impact on explicit memory, so someone driving a car on 'autopilot' may arrive at their destination safely, but with little memory of the journey. Thus, the concepts of automatic and controlled processing map closely onto the currently more fashionable concepts of unconscious or implicit processing and conscious or explicit processing. But this mapping raises another question: is consciousness something we use to control our behaviour, or do we *become* conscious of our behaviour when we exert control over it?

### 2.3 The neuropsychology of consciousness

The research presented in the previous sections raises questions about the nature of consciousness. Are we conscious of a stimulus (or of a memory) because it exceeds some threshold of salience or activation, or are we conscious of it because it is processed in a particular way? In other words, is consciousness something that might be associated with very many cognitive processes or is it a feature of particular cognitive modules? Is there even a unitary 'consciousness module'? There are no clear answers to these questions at present, but this section aims to show the potential for neuropsychology to help find answers. While you are reading it, bear in mind that the search for a consciousness module may be futile: Dennett (1991) argues that searching for a place where consciousness happens – a 'Cartesian theatre' – is a mistake based on a misunderstanding of consciousness. We shall meet Dennett again in Section 4.

Studies of altered consciousness following localized brain injury provide a way of assessing whether consciousness is a unitary or modular function. If it is a unitary function, it could be localized in a single 'consciousness area' of the brain, or distributed across a network of interconnected brain regions, or it could be the result of some non-localized process such as synchronized activity across brain regions

(see ffytche, 2000). If consciousness is a modular function, then our conscious awareness of colours might be caused by processes localized in quite different brain regions from, say, our conscious awareness of movements or sounds.

At first glance, neuropsychological studies suggest that consciousness is modular because brain injury often causes loss of consciousness of only a subset of sensations and cognitions. For example, people with amnesia are not conscious of information learned since the onset of their amnesia. People with unilateral neglect on the other hand have normal consciousness of their memories but lack consciousness of one side of space. Is it the case then that our consciousness for memories is separate from our consciousness for space, as a modular interpretation would suggest? The answer is not clear. Patients with amnesia may lack consciousness of their memories because they lack critical unconscious processes that feed into a unitary consciousness. Likewise neglect patients may not have suffered damage to a ‘consciousness of space’ module but rather have deficits in attentional processes that feed into a unitary consciousness.

Zeki and ffytche (1998) studied a blindsight patient known as G.Y. **Blindsight** is a disorder exhibited by some patients with brain damage leading to blindness in part of the visual field. In blindsight, there is a somewhat preserved ability to respond appropriately to visual stimuli in the blind region of the visual field, despite having no sense of seeing them. G.Y.’s blindness is selective, so that he reports having some sort of conscious experience of some visual stimuli in his blind field (e.g. fast-moving stimuli) but denies having any experience of other stimuli (e.g. slow-moving stimuli). Usually, G.Y. can identify a stimulus from a small selection of distractors only if he has some conscious experience of it. Occasionally, however, he can identify stimuli even without any conscious experience; that is, he exhibits blindsight for these stimuli.

Zeki and ffytche (1998) used functional magnetic resonance imaging (fMRI) to compare G.Y.’s brain activity with and without conscious perception of visual stimuli. G.Y. could detect the direction of movement of slow and fast-moving stimuli, but he was usually only conscious of the fast stimuli. The two types of stimuli differentially triggered activity in the motion cortex, the increase in activity being greater with the fast-moving stimuli for which G.Y. reported some conscious experience. Thus, consciousness of visual stimuli appears to be related to the amount of activity in a localized brain area, a brain area specialized for processing that type of stimulus rather than in a general ‘consciousness centre’. ffytche (2000) suggests that the increased local activity associated with consciousness may reflect the activity of additional populations of neurons in that brain region or it may reflect more complex processing by neurons that are also active when we are not conscious of the stimulus to which they respond. Note however that although Zeki and ffytche’s data support the hypothesis of a modular consciousness, our overall conscious experience – our individual ‘stream of consciousness’ – may still reflect the aggregation of processing across many brain regions. Consciousness for motion may be dissociable from consciousness for colour, for instance, but these two consciousness modules must somehow be bound together to produce our normal conscious experience of a moving coloured stimulus.

## Summary of Section 2

---

- Studies of priming or implicit memory show that people retain more information than they are aware of remembering.
- Studies of learning without awareness of what is learned have been hotly debated; so-called implicit learning might reflect failure to detect small amounts of awareness. A recent study of learning during anaesthesia suggests that memory priming occurs even when patients are unconscious.
- Schneider and Shiffrin (1977) argued that controlled search processes operate serially and are slow, increasingly so as task demands increase, because they depend on limited capacity short-term memory systems. Automatic search processes are fast and can operate in parallel because they operate on well-learned pathways in long-term memory. Automatic processing typically happens without awareness, whereas controlled processing is associated with conscious awareness of the task in hand and explicit memory for its products.
- Neuropsychological studies can help determine the structure of consciousness. ffytche (2000) uses a study of blindsight to argue that consciousness is a function of modular brain systems. Alternative arguments are that there is a single 'consciousness module' or that consciousness is a distributed function.

## 3 What is consciousness for?

### 3.1 Consciousness and behavioural control

The studies of implicit learning and memory discussed in the previous section suggest that it is possible to learn about a variety of different stimuli without being conscious of them. We appear to 'remember' things without any conscious experience of doing so. If we can do this without consciousness, does consciousness actually serve any function or is it an epiphenomenon, a by-product of brain processes that does not in itself affect the system? This section discusses evidence that consciousness serves a variety of purposes that together make us able to function effectively even in novel environments.

Research into automatic and controlled processes shows that automatic or unconscious processes tend to be fast and efficient but inflexible. Providing there is no competition for sensory systems (e.g. trying to view two complex pictures simultaneously) or response effectors (e.g. trying to write two answers simultaneously), several automatic processes can run concurrently. Controlled or conscious processes, on the other hand, are slower and more demanding of cognitive resources, so it is hard to carry out more than one at a time. Stimuli must therefore be selected for conscious processing. The Cheshire cat illusion, discussed in Activity 15.4 in Section 3.2, illustrates this point. Chapter 2, on attention, discusses how this selection is done. Despite their disadvantages, we need conscious processes when performing new tasks or trying to override habits. This is illustrated by the tendency

of people to generate stereotyped responses when distracted from a random generation task, a task that requires frequent strategy shifts to avoid lapsing into stereotyped response patterns. Baddeley *et al.* (1998) asked participants to generate random sequences of digits or key presses. They manipulated the availability of controlled processing resources by asking participants to perform other tasks at the same time (for example, solving problems or retrieving information from long-term memory). In these conditions, faster and more automatic processes generated stereotyped responses such as '1, 2, 3' or parts of familiar telephone numbers. Conscious or controlled processes thus seem to be associated with flexible responding. Box 15.2 on affective priming suggests that conscious processes help us make rational rather than emotional decisions. The research presented in Box 15.3 on errorless learning shows that making mistakes prevents people with amnesia from learning as effectively as they might. With normal memory, consciousness of our memory for past errors may help us adapt our behaviour by learning from our mistakes.

## 15.2

## Research study

### Affective priming

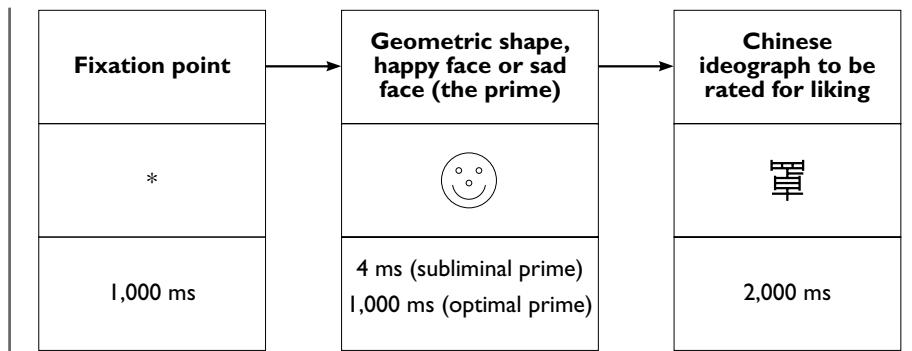
Zajonc (1980) hypothesized that emotional or affective responses can be triggered by information that undergoes only minimal processing. Cognitive responses require more processing. This is known as the 'affective primacy hypothesis'. Murphy and Zajonc (1993) tested this hypothesis in a series of experiments that compared the influence of subliminal and 'optimal' (i.e. consciously visible or supraliminal) primes on responses to subsequent stimuli.

In the first experiment in their study, Murphy and Zajonc asked 32 participants to rate their liking of Chinese ideographs, on a scale of 1 = 'did not like the ideograph at all' to 5 = 'liked the ideograph quite a bit'. Each ideograph was shown for 2 seconds. There were four types of trial:

- *no-prime controls*, where the ideographs were shown alone
- *irrelevant prime controls*, where a geometric shape preceded each ideograph
- *positive affective prime trials*, where a photograph of a happy face preceded each ideograph
- *negative affective prime trials*, where a photograph of an angry face preceded each ideograph.

For participants in the subliminal prime condition, each slide of a shape or face was presented for just 4 ms, followed immediately by an ideograph that served both as the stimulus to be rated and as a visual mask to prevent the image of the prime lingering in iconic memory. Participants in the optimal prime condition viewed each prime slide for 1,000 ms, followed immediately by the ideograph. A summary of these experimental trials is shown in Figure 15.3.





**Figure 15.3** Experimental trials used to demonstrate affective priming

The results are shown in Table 15.1. The ideographs were rated as more pleasant when preceded by a subliminal positive prime, compared with the no prime and irrelevant prime control conditions, and more unpleasant when preceded by a subliminal negative prime. In contrast, the visible or ‘optimal’ primes had no effect on ratings of the ideographs.

**Table 15.1** Mean ratings of liking for Chinese ideographs in affective prime, irrelevant prime and no prime conditions

Condition	Subliminal presentation (4 ms)	Optimal presentation (1,000 ms)
Positive prime	3.46	3.02
Negative prime	2.70	3.28
Irrelevant prime	3.06	3.15
No prime	3.06	3.11

In a follow-up experiment, Murphy and Zajonc tested the effects of subliminal and optimal primes on ratings of the size of object represented by each ideograph. This time the relevant primes were small or large shapes and the irrelevant primes were faces with neutral expressions. The results contrasted with those of the experiment described above: the subliminal primes had no effect on ratings of size whereas the large optimal primes led to higher ratings and the small optimal primes to lower ratings.

#### Conclusion

Murphy and Zajonc (1993) argued that the affective primes altered participants’ mood, in the sense that participants had an emotional response to the primes. When participants were aware of their mood change and of its source, they could ignore it when rating the ideographs. However, when they were unaware of it, their affective response to the primes ‘spilled over’ onto the rating task. For negative primes, the authors referred to this effect as ‘free floating anxiety’ – that is, anxiety without awareness of what caused it or what we are anxious about. Consciousness of the primes allowed participants to override their affective response to them when judging the ideographs.

**15.3****Research study****Errorless learning**

Baddeley and Wilson (1994) argued that one of the main functions of explicit memory is to help us learn from our mistakes. Without awareness of our errors, past mistakes serve only to prime similar mistakes in the future. They tested their hypothesis by comparing two modes of learning in 32 participants with normal memory (16 young and 16 elderly adults) and 16 participants with amnesia who were assumed to have normal implicit memory combined with impaired explicit memory. Participants with normal memory received one list of 10 words in the errorful learning condition and another list of 10 words in the errorless learning condition. Amnesic participants received five words in each condition, to avoid floor and ceiling effects. The stimulus words were all five letters long and chosen because their two-letter stems could be completed in several ways. For example, the stem QU– could be completed as QUOTE (the stimulus in this case) or QUIET, QUEEN, QUACK, etc. The large number of potential completions maximized the possibility for making errors in the errorful learning condition.

In the errorful learning condition, participants were told that the experimenter was thinking of a five-letter word beginning QU– and asked to guess what the word might be. After making up to four incorrect guesses, the participant was told that the word was QUOTE (or a back-up word if they happened to guess the target straight away) and asked to write it down. This procedure was repeated for the other words in the list, and then again two more times for the entire list.

In the errorless learning condition, participants were told 'I am thinking of a five-letter word beginning with QU and that word is QUOTE please write that down.' The list of target words was presented three times, as in the errorful condition.

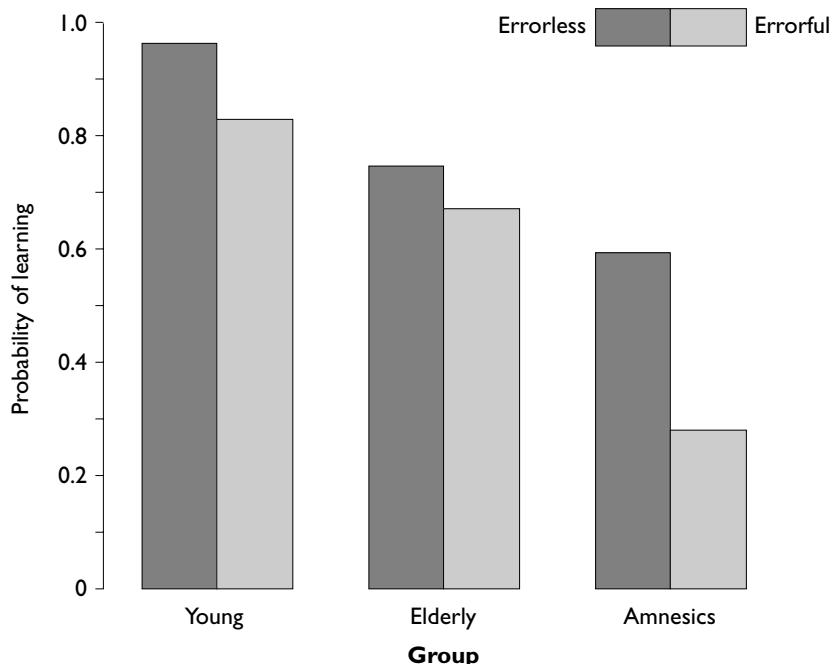
These first three trials were termed the pre-training phase. Learning condition was only manipulated during this pre-training phase. The test phase comprised nine further learning trials. On each of these trials, regardless of the initial learning condition, the experimenter provided the first two letters of each word in the list and asked the participant to write down a word starting with those letters from the earlier list. If they could not remember a word from the previous phase, they were asked to say any word that came to mind beginning with those letters. In the case of incorrect responses, the experimenter provided the correct word. Baddeley and Wilson analysed performance in the test phase in terms of the probability of learning – that is, the probability of an item that is not known on one trial becoming learned on the next trial. The learning probabilities for the two learning conditions are illustrated in Figure 15.4 overleaf.

**Conclusion**

Participants with amnesia benefited considerably more from the errorless learning procedure than the young and elderly participants with normal memory. Without explicit memory for their errors, amnesic participants were unable to correct their mistakes on subsequent trials and so found the learning task



particularly difficult when they were encouraged to make errors on the initial, pre-training trials. Errorless learning techniques have been used to teach amnesics useful information such as how to program a personal organizer to remind them of appointments.



**Figure 15.4** Learning probabilities for young, elderly and amnesic participants' learning conditions

Source: Baddeley and Wilson, 1994, Figure 3, p.59

### 3.2 Cross-talk between cognitive modules

An important feature of consciousness is that it seems to break the modularity of mind. Cognitive psychologists often assume that cognitive processes are modular; that is, that they operate in clusters that function independently from other clusters of processes (see Section 3 in Chapter 17). Each cluster or ‘module’ processes a particular sort of information, so there may be one module for comprehending spoken language and another module for recognizing faces. Consciousness involves cross-talk between these otherwise independent modules. If we are conscious of something, we can talk about it, decide to touch it or ignore it, imagine holding or owning it, have beliefs about it. In studies of blindsight, note that patients do not tend to initiate responses towards stimuli. Even though they can respond appropriately to stimuli when encouraged to guess, their residual processing of the stimuli is of little practical use to them. Without the conscious experience of seeing, someone with this condition would not, for example, pick up a glass of water placed in their blind field even if they were thirsty and could point to its location when encouraged to do so. Normally, seeing a glass of water means that we can also drink it or talk about it – consciousness of the visual percept makes it available to action and language

modules. In blindsight, although some visual information is processed and influences ‘guessing’ behaviour, that information is not sufficient to break out of the visual perception module to form a basis for conscious behaviour. Likewise with implicit learning. Although exposure to covert regularities may improve performance on an implicit learning task with similar stimuli, even without awareness of what has been learned, there is limited transfer of this improvement to tasks with different stimuli constructed according to the same rules (e.g. Gomez, 1997). When we are aware of what we have learned, we are better able to apply that learning to novel tasks.

Although cognitive research has helped identify the functions of consciousness, for controlling our behaviour and allowing cross-talk between cognitive modules, it does not explain why these functions are associated with conscious experience. Could it be possible, for a computer, say, to perform these functions without consciousness? Philosophers use the term ‘zombie’ to refer to the idea of someone exactly like us, with the same cognitive processes as us, the same knowledge, memories, planning abilities, etc., but without consciousness. If this idea makes sense to us (even though zombies are a fiction), then it suggests that the ‘hard problem’ of consciousness is indeed a very hard problem because there is nothing about our cognitive processes that necessitates the conscious experience that accompanies them (see Chalmers, 1996).

### 3.3 Altered states of consciousness

The cognitive basis of altered states of consciousness is not well understood. Nonetheless, altered states are interesting because, by providing a contrast, they help us reflect on what ‘normal’ consciousness is like. Because they are generally dysfunctional states, in the sense that they are not conducive to normal everyday behaviour, they give us clues about the functions of normal conscious states. Altered states are therefore included briefly in this chapter as a discussion point to help you think about the possible functions of consciousness.

Drugs such as ketamine and lysergic acid diethylamide (LSD) cause hallucinations and other perceptual disturbances such as synaesthesia, a condition where stimuli in one sensory modality trigger experiences in another sensory modality (e.g. touching something hard may produce the sensation of seeing green). It appears that these drugs cause a flooding of the sensory system and a breaking down of the modularity of sensory systems. Normal consciousness may therefore involve selecting incoming sensory information to prevent too much information reaching higher-level cognitive processes. Drugs such as alcohol cause loss of inhibition, making us more likely to say things or do things that we would refrain from doing in our normal conscious state. Normal consciousness may therefore involve monitoring and controlling our behaviour.

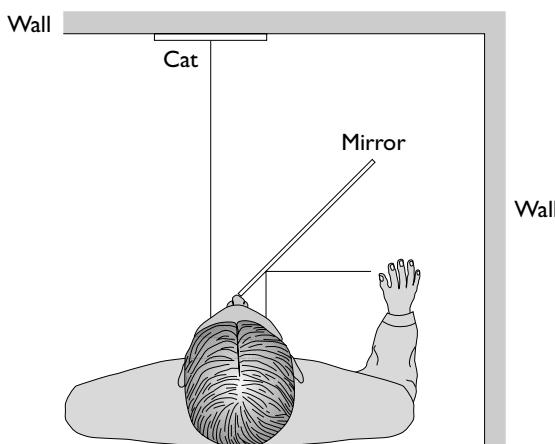
Hypnosis is a state of deep relaxation that makes people more susceptible to suggestion. There is debate about whether it is truly an altered state of consciousness, or merely a response to the particular combination of relaxation and social pressure to conform to the hypnotist’s suggestions. Nonetheless, hypnotized people can perform surprising feats, such as speaking ‘forgotten’ languages from their childhood or even undergoing minor surgery without painkillers or anaesthetic. One explanation of such feats is that hypnosis reduces

our normal tendency to check our mental contents against the outside world. This checking and updating of our mental model of ourselves and our environment is called **reality monitoring**. With reduced reality monitoring, hypnotized people become more credulous because they are less likely to check the hypnotist's suggestions against what they know to be true. This makes them better able to maintain a 'hallucination' even if it contradicts incoming sensory information.

Altered states of consciousness suggest that normal consciousness operates with or on a selected portion of the information that constantly bombards our senses. For a demonstration of this selection, try the Cheshire cat activity (Activity 15.4). Normal consciousness also involves checking our current mental state against incoming information from our environment and checking our behaviour against our intended goals.

#### ACTIVITY 15.4

Sit facing a picture of a cat (or a real one if you can persuade it to sit still) on an otherwise blank wall, with another blank wall to your right (see Figure 15.5). Hold the edge of a mirror against your nose, and tilt it so you view the cat picture with your left eye only. Hold up your right hand so that it is reflected in the mirror, the reflection being viewed by your right eye. Move your right hand slowly towards and away from you. With appropriate adjustments to the mirror or to the direction of hand waving, you should experience the Cheshire cat illusion: moving your hand appears to 'rub out' parts or all of the cat (Duensing and Miller, 1979).



**Figure 15.5** Set-up for the Cheshire cat illusion

When you experience the Cheshire cat illusion, your brain is receiving two separate visual images, yet you only see one of them at a time. What does this tell you about consciousness?

**Note:** If you have difficulty seeing this illusion, try this simpler version. Roll a sheet of paper into a tube and look through it with your left eye while viewing your right hand with your right eye. Hold your right hand against the tube, about 10 cm from your eye. Move your hand slightly to and fro until you perceive a hole in your hand with the same diameter as the tube.

## COMMENT

The illusion occurs because the two percepts – the one of the cat and the one of your hand – cannot be fused. Only one percept can be conscious at a time, hence if you are conscious of seeing the reflection of your hand, you will not be conscious of seeing the part of the cat picture that occupies the same apparent visual location as the reflection of your hand. This illusion illustrates the selective and limited capacity aspect of consciousness. The phenomenon is known as ‘binocular rivalry’ and has been used by neuroscientists to determine the neuronal activity associated with consciousness of a visual stimulus (e.g. Logothetis and Schall, 1989).

---

## Summary of Section 3

- Consciousness is associated with, and may cause, behavioural control. It helps us override habitual and emotional responses and to learn from our mistakes.
- The effects of mind-altering drugs, and of hypnosis, suggest that normal conscious states involve monitoring and control of behaviour, and selection of incoming information. We are only conscious of a small amount of information at any time.

## 4 Cognitive theories of consciousness

This section focuses on two theories: working memory and global workspace theory. Baddeley’s working memory model is chosen because it is already a widely used model in cognitive psychology (see Chapter 9 on working memory). If it can also say something about consciousness then it could help us integrate our thinking about consciousness with our existing understanding of other aspects of cognition. Baars’s global workspace approach is discussed as a contrast, partly because it is the most detailed and comprehensive cognitive model of consciousness currently available and partly because Baars explicitly discusses the hypothetical relationship between consciousness and working memory.

Short-term memory, and later working memory, have often been identified with consciousness. Thus James (1918) described memory generally as a way of bringing back past conscious experiences, but argued that for primary memory there is nothing to bring back because ‘it was never lost; its date was never cut off in consciousness from that of the immediately present moment’ (1918, p.647). Baddeley (1993) portrays working memory as a conduit to consciousness that serves to bring together information in different modalities from perception and long-term memory, enabling us to imagine novel solutions to problems of evolutionary significance. For example, he suggests that a vivid image of a hunting ground, which included locations where food was found before and locations where danger lurked, would have evolutionary significance as a tool for predicting events and planning action. Working memory thus presents one solution to the binding problem (see Chapter 9, Section 2.3.4), the problem of creating a coherent, unified conscious

experience from disparate sensory inputs. (Note though that the binding problem also pertains at lower levels of cognition; we also need to bind together the results of independent shape, colour, motion and location processes, for example). Baddeley (2000) proposed a new component to the working memory model, called the episodic buffer, as a temporary store for bound information and as an interface between working memory, long-term memory and consciousness. However, the revised model does not specify whether information is conscious by virtue of being stored in the episodic buffer or whether it is conscious only when acted upon by central executive processes. Nonetheless, the wealth of evidence that working memory is involved in conscious cognition, in problem solving and mental imagery for example, suggests that working memory must actually be a cognitive correlate of consciousness.

The hypothetical central executive (see Chapter 9, Section 2.1) of working memory seems particularly closely correlated with consciousness, playing a key role in conscious, strategic activities such as retrieval of information from long-term memory, selective attention, and so on (Baddeley, 1996). Baddeley (1986) used Norman and Shallice's (1986) cognitive model of contention scheduling and a supervisory attentional system (SAS) as a model of the central executive. Norman and Shallice's model aimed to explain action selection. They hypothesized that routine responses are selected by a relatively automatic process called contention scheduling, essentially the selection of habitual action schemata by virtue of their high activation level and ability to inhibit competing schemata. Changes in behaviour are effected by an SAS system that increases the activation of a non-habitual action schema so that it 'wins' in the contention scheduling process. The SAS thus serves as an error-correcting mechanism and as the basis for conscious behavioural control, allowing us to respond appropriately in novel situations. Errors in behaviour, such as everyday action slips (e.g. adding milk from habit when intending to make a cup of black coffee) and perseverative behaviours following frontal brain injury (i.e. persisting with an established response when altered conditions mean a new response is needed) are explained as failures of the SAS. Baddeley's adoption of the SAS as a model of the central executive extended its remit from the conscious control of action to the conscious control of cognition generally. However, both models, the SAS and the central executive, are subject to the criticism of postulating little more than a 'homunculus', a little person in the head that tells us which action to do or which memory to retrieve.

Dennett (1991) refers to the homunculus problem as the problem of the Cartesian theatre: it arises from the assumption that there is a cognitive or neuroanatomical module that 'does' consciousness, a location in the brain where consciousness occurs. Dennett's **multiple drafts** theory of consciousness offers a way out of this problem. Dennett argues that it is misleading to think of consciousness as something that suddenly happens, in the sense that stimulus processing works its way up from low-level sensory processes to higher-level cognitive processes and somewhere along the way our representation of the stimulus suddenly enters consciousness. In Dennett's theory, stimuli are not processed and then sent to a consciousness module, they are just processed. Which of many parallel streams of processing we become aware of, and how we experience them, depends on how and when the system is 'probed' by tasks that require particular responses.

Whereas Baddeley argues that working memory is necessary for consciousness, Baars (2002) argues that consciousness is necessary for working memory. All elements of ‘active working memory’, such as subvocal rehearsal and visual imagery, are conscious. Baars describes consciousness as a **global workspace**, a means of bringing together the products of processing from widely distributed modules. This bringing together – of inner speech, mental imagery, strategic recall, etc. – is necessary for working memory to function. Voluntary control of behaviour is also dependent on consciousness, requiring conscious goals and conscious perception of the effects of our actions. Thus we can learn from our mistakes or avoid emotions biasing our behaviour because our behavioural control processes have access to our knowledge of past mistakes or the causes of our mood swings.

Baars (1997) uses the analogy of a theatre. The unconscious processes of syntax analysis, visual boundary analysis, semantic processing, etc. are the stagehands working behind the scenes in the theatre of consciousness. Although there are actors on the stage, we only see the actor currently performing in the spotlight. Working memory forms the stage of consciousness, representations in working memory are the actors on the stage. Actors only step into the spotlight when chosen by the stage director – likewise the contents of working memory have the potential to become conscious but usually do not do so unless selected by the central executive. Once in the spotlight, the actor can be seen by everyone else: once a representation becomes conscious, it is accessible to other cognitive processes. Thus, consciousness overcomes some of the modularity of mind, enabling us to talk about our ideas, express our feelings, use remembered information to solve problems, and so on.

Baars (2002) draws on many recent neuroscience studies to support his theory. The essence of these findings is that unconscious processing of stimuli activates localized brain regions whereas conscious processing of the same stimuli activates widely distributed brain regions. For example, an fMRI study by Dehaene *et al.* (2001) showed that processing of masked visual words was associated with activation in early visual cortex whereas processing of visible, unmasked words was also associated with activation in parietal and prefrontal cortex. However, the finding by Zeki and ffytche (1998), of increased but still localized activation with conscious perception, seems to contradict Baars’s theory (see Section 2.3). Thus it appears that consciousness is often but not always associated with global brain activity.

## Summary of Section 4

---

- Working memory is closely allied to consciousness, providing a way of binding together information from perception and long-term memory to create conscious, multi-modal representations.
- Although working memory is a cognitive correlate of consciousness, the relationship between working memory and consciousness is unclear. Baars (1997) suggested that working memory serves to select the information that will become conscious. More recently, he has argued that consciousness is necessary for working memory (Baars, 2002).

- Baars sees consciousness as a global workspace, enabling cognitive modules to share information. ‘Global access’ to conscious information is consistent with findings of widespread brain activity when participants are aware of a stimulus, contrasted with more localized activity when they are unaware of it.

## 5 Conclusion: what can cognitive psychology tell us about consciousness?

Cognitive psychology has helped to generate hypotheses about the functions of consciousness. For example, studies of errorless learning and of priming have suggested that consciousness helps us control our behaviour by avoiding repetitions of our mistakes and suppressing emotional, rather than rational, responses. However, although consciousness is associated with behavioural control, we have no evidence that consciousness causes behavioural control. Similarly, we have seen that consciousness is associated with limited resource systems like working memory and selective attention, but we do not know whether consciousness plays a causal role in remembering or attending. Thus cognitive psychology has helped us to discover the cognitive correlates of consciousness – that is, those cognitive processes that are always accompanied by consciousness. But a correlation only tells us that two things are related, not that one causes the other. This problem is exemplified by the debate outlined in Section 4 about whether working memory is necessary for consciousness or vice versa. Hardcastle (2000) discusses this problem of correlational data in relation to the search for the neural correlates of consciousness. She argues that there will be many correlates of consciousness, but to explain consciousness effectively we must identify the ‘proximal cause’, the correlate that is the most important for consciousness. She uses the example of depression: is Fred depressed because he has received bad news or because the level of noradrenaline in his brain has dropped? If people are more likely to become depressed as the result of neurotransmitter changes than as the result of hearing bad news, then neurotransmitter levels are the more important correlate of depression. Hardcastle suggests that it is too early to be able to say which of the many correlates identified so far is the proximal cause of consciousness. Indeed, it is not even clear whether the answer will lie with those who seek to reduce consciousness to particular brain modules or individual neural events, or with those who argue for a dynamic systems approach, who would suggest that consciousness emerges from complex and global interactions of brain processes. In the meantime, cognitive psychology provides us with useful techniques for analysing and researching consciousness. At least, it provides possible solutions to the easy problems of consciousness – attention, recollection, voluntary action and so forth. The problem of why consciousness feels the way it does remains a very hard problem.

## Further reading

Baars, B. (1988) *A Cognitive Theory of Consciousness*, New York, Cambridge University Press, and Baars, B. (1997) *In the Theater of Consciousness: The*

*Workspace of the Mind*, New York, Oxford University Press. These books explain Baars's global workspace theory of consciousness.

Blackmore, S. (2001) 'Consciousness', *The Psychologist*, vol.14, no.10, pp.522–5. A succinct overview of interesting issues.

Blackmore, S. (2004) *Consciousness: An Introduction*, Oxford, Oxford University Press. A comprehensive and accessible introduction to philosophical and empirical issues in consciousness studies. Includes interesting practical exercises to help you think about key problems.

Metzinger, T. (ed.) (2000) *Neural Correlates of Consciousness*, Cambridge, MA, MIT Press. An excellent collection of original chapters by leading consciousness researchers. Particularly recommended are the chapters by Hardcastle and ffytche.

Young, A.W. and Block, N. (1996) 'Consciousness', in Bruce, V. (ed.) *Unsolved Mysteries of the Mind*, Hove, Erlbaum. Useful discussion of access and phenomenal consciousness and of what consciousness researchers can learn from neuropsychology.

## References

- Andrade, J. (1995) 'Learning during anaesthesia: a review', *British Journal of Psychology*, vol.86, no.4, pp.479–506.
- Atkinson, R.C. and Shiffrin, R.M. (1968) 'Human memory: a proposed system and control processes', in Spence, K.W. and Spence, J.D. (eds) *The Psychology of Learning and Motivation* (vol.2), New York, Academic Press.
- Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, New York, Cambridge University Press.
- Baars, B.J. (1997) *In the Theater of Consciousness: The Workspace of the Mind*, New York, Oxford University Press.
- Baars, B.J. (2002) 'The conscious access hypothesis: origins and recent evidence', *Trends in Cognitive Sciences*, vol.6, no.1, p.47.
- Baddeley, A.D. (1986) *Working Memory*, Oxford, Oxford University Press.
- Baddeley, A.D. (1990) *Human Memory: Theory and Practice*, Hove, Lawrence Erlbaum Associates.
- Baddeley, A.D. (1993) 'Working memory and conscious awareness', in Collins, A.F., Gathercole, S.E., Conway, M.A. and Morris, P.E. (eds) *Theories of Memory*, Hove, Lawrence Erlbaum Associates.
- Baddeley, A.D. (1996) 'Exploring the central executive', *Quarterly Journal of Experimental Psychology*, vol.49, pp.5–28.
- Baddeley, A.D. (2000) 'The episodic buffer: a new component of working memory?', *Trends in Cognitive Science*, vol.4, no.11, pp.417–23.
- Baddeley, A.D., Emslie, H., Kolodny, J. and Duncan, J. (1998) 'Random generation and the executive control of working memory', *Quarterly Journal of Experimental Psychology*, vol.51A, no.4, pp.819–52.
- Baddeley, A.D. and Wilson, B.A. (1994) 'When implicit learning fails: amnesia and the problem of error elimination', *Neuropsychologia*, vol.32, pp.53–68.

- Banks, W.P. (1993) 'Problems in the scientific pursuit of consciousness', *Consciousness and Cognition*, vol.2, no.4, pp.255–63.
- Bargh, J.A., Chen, M. and Burrows, L. (1996) 'Automaticity of social behavior: direct effects of trait construct and stereotype activation on action', *Journal of Personality and Social Psychology*, vol.71, pp.230–44.
- Block, N. (1995) 'On a confusion of a function of consciousness', *Behavioral and Brain Sciences*, vol.18, no.2, pp.227–87.
- Chalmers, D. (1996) *The Conscious Mind*, Oxford, Oxford University Press.
- Deeprose, C., Andrade, J., Varma, S. and Edwards, N. (2004) 'Unconscious learning during surgery with propofol anaesthesia', *British Journal of Anaesthesia*, vol.92, pp.171–7.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.-F., Poline, J.-B. and Rivière, D. (2001) 'Cerebral mechanisms of word masking and unconscious repetition priming', *Nature Neuroscience*, vol.4, pp.752–8.
- Dennett, D. (1991) *Consciousness Explained*, Boston, MA, Little, Brown & Co.
- Duensing, S. and Miller, B. (1979) 'The Cheshire Cat effect', *Perception*, vol.8, pp.269–73.
- Eich, E. (1984) 'Memory for unattended events: remembering with and without awareness', *Memory and Cognition*, vol.12, pp.105–11.
- ffytche, D (2000) 'Imaging conscious vision', in Metzinger, T. (ed.).
- Gomez, R.L. (1997) 'Transfer and complexity in artificial grammar learning', *Cognitive Psychology*, vol.33, no.2, pp.154–207.
- Hardcastle, V.G. (2000) 'How to understand the N in NCC', in Metzinger, T. (ed.).
- Jacoby, L.L., Woloshyn, V. and Kelley, C. (1989) 'Becoming famous without being recognized: unconscious influences of memory produced by dividing attention', *Journal of Experimental Psychology: General*, vol.118, no.2, pp.115–25.
- James, W. (1918, first published in 1890) *The Principles of Psychology*, vol.1, London, Macmillan and Co. Ltd.
- Knowlton, B.J., Ramus, S.J. and Squire, L.R. (1992) 'Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances', *Psychological Science*, vol.3, no.3, pp.172–9.
- Kunst-Wilson, W.R. and Zajonc, R.B. (1980) 'Affective discrimination of stimuli that cannot be recognized', *Science*, vol.207, pp.557–8.
- Levine, J. (1983) 'Materialism and qualia: the explanatory gap', *Pacific Philosophical Quarterly*, vol.64, pp.354–61.
- Lieberman, M.D. (2000) 'Intuition: a social cognitive neuroscience approach', *Psychological Bulletin*, vol.126, no.1, pp.109–37.
- Logothetis, N.K. and Schall, J.D. (1989) 'Neuronal correlates of subjective visual perception', *Science*, vol.245, pp.761–3.
- Marcel, A.J. (1983) 'Conscious and unconscious perception: experiments on visual masking and word recognition', *Cognitive Psychology*, vol.15, pp.197–237.
- Marcel, A.J. (1988) 'Phenomenal experience and functionalism', in Marcel, A.J. and Bisiach, E. *Consciousness in Contemporary Science*, Oxford, Oxford University Press.

- McElree, B. (2001) ‘Working memory and focal attention’, *Journal of Experimental Psychology: Learning, Memory, and Perception*, vol.27, no.3, pp.817–35.
- Metzinger, T. (ed.) (2000) *Neural Correlates of Consciousness*, Cambridge, MA, MIT Press.
- Meyer, D.E. and Schvaneveldt, R.W. (1971) ‘Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations’, *Journal of Experimental Psychology*, vol.90, pp.227–35.
- Moray, N. (1959) ‘Attention in dichotic listening: affective cues and the influence of instructions’, *Quarterly Journal of Experimental Psychology*, vol.11, pp.56–60.
- Murphy, S.T. and Zajonc, R.B. (1993) ‘Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures’, *Journal of Personality and Social Psychology*, vol.64, pp.723–39.
- Neumann, R. and Strack, F. (2000) “‘Mood contagion’ – the automatic transfer of mood between persons”, *Journal of Personality and Social Psychology*, vol.79, no.2, pp.211–23.
- Nissen, M.J. and Bullemer, P. (1987) ‘Attentional requirements of learning: evidence from performance measures’, *Cognitive Psychology*, vol.19, pp.1–32.
- Norman, D.A and Shallice, T. (1986) ‘Attention to action: willed and automatic control of behavior’, in Davidson, R.J., Schwartz, G.E. and Shapiro, D. (eds) *Consciousness and Self-Regulation*, New York, Plenum.
- Reber, A.S. (1967) ‘Implicit learning of artificial grammars’, *Journal of Verbal Learning and Verbal Behavior*, vol.6, pp.855–63.
- Schneider, W. and Shiffrin, R.M. (1977) ‘Controlled and automatic human information processing: 1. Detection, search, and attention’, *Psychological Review*, vol.84, pp.1–66.
- Shanks, D.R. and St John, M.F. (1994) ‘Characteristics of dissociable human learning systems’, *Behavioural and Brain Sciences*, vol.17, no.3, pp.367–95.
- Squire, L.R. and McKee, R. (1992) ‘Influence of prior events on cognitive judgments in amnesia’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.18, no.1, pp.106–15.
- Zajonc, R.B. (1980) ‘Feeling and thinking: preferences need no inferences’, *American Psychologist*, vol.35, pp.151–75.
- Zeki, S. and ffytche, D.H. (1998) ‘The Riddoch syndrome: insights into the neurobiology of conscious vision’, *Brain*, vol.121, pp.25–45.



# Cognitive modelling and cognitive architectures

Chapter 16

Paul Mulholland and Stuart Watt

---

## 1 What is cognitive modelling?

In this chapter, we are going to return to a psychological method that was introduced in Chapter 1 and has been raised at various points since: cognitive modelling. You have already met a few cognitive models, although you may not have realized that they were cognitive models at the time. Now is a good time to look at the method of cognitive modelling in a bit more detail and see the contributions that models make to cognitive psychology.

First of all, what is cognitive modelling? **Cognitive modelling** involves building a working model of a cognitive process and then comparing the behaviour of that model against human performance. If the model behaves in the same way as humans, then the structure of the model and the way it works may give some insight into how humans perform the task. Cognitive modelling has been used to help explain a range of cognitive processes such as face recognition (see Chapter 4), language comprehension (Chapter 6) and analogical reasoning (Chapter 10).

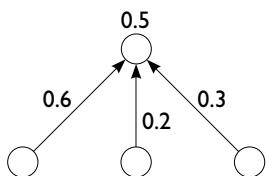
Cognitive modelling is very like the kind of technique that car designers or architects use to make it easier to see and test their designs. Psychologists use models in the same way: to make it easier to understand and to test their theories. Generally speaking, there are two different approaches to cognitive modelling. First, there is a high-level approach. To follow the car analogy, a high-level model should look and behave as much like a car as possible, without necessarily having the same internal workings. A high-level model might state that there is an engine, but might not say exactly how it worked, or what it was made of. But there is also a low-level approach, where a modeller would look at representing the kinds of bits that cars were made from (wheels, axles, pistons, valves, and so on) and try to understand how the behaviours of these components could work together to behave like a car. In cognitive psychology, Parallel distributed processing (or connectionism) can be thought of as a low-level modelling approach and rule-based systems can be thought of as a high-level approach. These will be considered in turn in this section.

### 1.1 Parallel distributed processing

**Parallel distributed processing** or PDP modelling (Rumelhart and McClelland, 1986), sometimes known as connectionist or neural network modelling, involves building models that match human performance by programming artificial neurons into networks. The artificial neurons with which the model is built, though far simpler than the actual neurons found in the human brain, have certain neural-like

properties (such as the way activation can spread between them). Also, the structure of these networks, in terms of the number of artificial neurons and their interconnections, is far simpler than the human brain. However, a PDP model can provide many useful insights into how a neural-like architecture can exhibit human-like cognitive behaviour.

The artificial neurons that make up a PDP model are often referred to as **units** or **nodes**. These units are connected together by links to form a network. A very simple network of four units is shown in Figure 16.1. Here, three units are providing input to a fourth unit. PDP units can be activated and can then spread their activation along the specified links to subsequent units. In this simple example, the three ‘sending’ units are assumed to each have an activation level of +1. The strength of the signal received by the next unit is determined by the products of the sending units’ activation levels (+1) and the **weight** of the link to the (fourth) ‘receiving’ unit. In Figure 16.1, each of the links has been assigned a numerical weight (usually a value between  $-1$  and  $+1$ ): the left-hand link has been assigned a weight of 0.6; the middle link has a weight of 0.2; the right-hand link has a weight of 0.3.



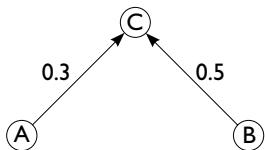
**Figure 16.1** A network of four nodes

Whether the input to a unit is sufficient to activate it is determined by the unit’s **threshold value**. The threshold value is the level of input that a unit must receive in order to send an output to subsequent units in the network. For example, in Figure 16.1, the top unit receiving input from the other three has a threshold value of 0.5. The threshold value is shown above the unit. (For simplicity, threshold values have not been specified for the lower three units.) As the top unit has a threshold value of 0.5, this means that an input from the left-hand link, having an activation of 1.0 and a weight of 0.6, would itself be sufficient to activate the unit (since  $1.0 \times 0.6 = 0.6$ , which is greater than the threshold of 0.5). The middle and right-hand links *individually* would not activate the unit, as the sum of the inputs from these units ( $1.0 \times 0.2 = 0.2$  and  $1.0 \times 0.3 = 0.3$  respectively) are lower than the threshold value. However, these two links working simultaneously can activate the unit as 0.2 and 0.3 add up to the threshold value of 0.5.

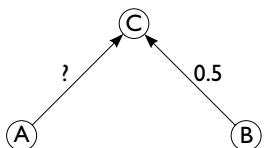
### ACTIVITY 16.1

---

- Assuming that units A and B are activated simultaneously, and both have an activation level of 1.0, what is the maximum threshold value of unit C that would still allow it to be activated?

**Figure 16.2**

- (b) Assuming that units and A and B are activated simultaneously, and both have an activation level of 1.0, what is the minimum weight of the link from unit A that would allow unit C to be activated if it had a threshold value of 0.6?

**Figure 16.3****COMMENT**

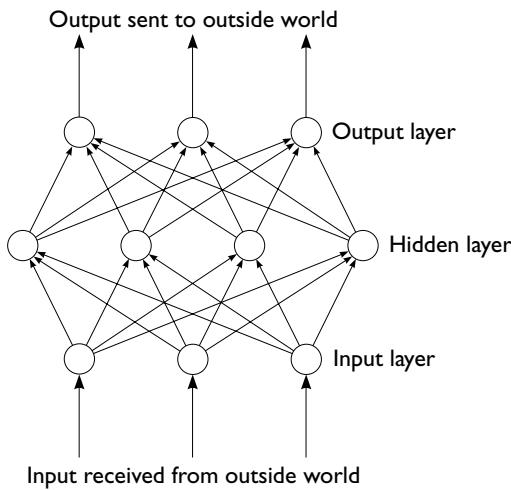
- (a) The maximum threshold value would be 0.8. This would be the sum of the two inputs  $1.0 \times 0.3 = 0.3$  and  $1.0 \times 0.5 = 0.5$ . If the threshold value was higher than 0.8 it would be impossible for the unit to fire given these two inputs.
- (b) The minimum weight would be 0.1. This would make the sum of the two inputs 0.6, equal to the threshold value of unit C. If the weight of the link was lower than 0.1 it would be impossible for unit C to fire given a threshold value of 0.6.

Note that the computation of inputs, outputs and activation values is usually considerably more complex than this example suggests. Nonetheless, the example does illustrate how such constructs interact in determining the behaviour of the network.

---

One of the most common forms of PDP network is a feed-forward, three-layer network, consisting of an input layer, hidden layer and output layer. An example is shown in Figure 16.4 overleaf. The **input layer** of the network contains the units that receive input from the outside world. If activated, these units send outputs to nodes in the **hidden layer**, which has no direct link to the outside world. Finally, the units of the **output layer** receive input from the hidden layer and send an output to the outside world. Whether the units of each layer send a signal to the next layer is determined by the inputs they receive, the weights of the links by which they receive their inputs, and the threshold value of the unit. The behaviour of a network (i.e. the outputs it provides according to inputs received from the outside world) will be determined by the threshold values of the units and the weights of the links.

PDP models can also ‘learn’ – one way in which this can be done is via an automatic procedure for successively modifying the weights of links until the network ultimately produces the correct outputs. During **training**, the network will be provided with a large number of example inputs, and a specification of the output the network should produce for each one. For example, suppose we present a



**Figure 16.4** A PDP architecture comprising an input layer, hidden layer and output layer

Source: adapted from Rumelhart and McClelland, 1986

network with an input pattern and a unit in the output layer provides an output when it is not desired. This can be corrected in the model by decreasing the weights of the links that provide this unit's input, and so making it more difficult for the unit to reach its threshold value when the network is presented with the same input pattern again. Conversely, if a unit in the output layer fails to provide an output when desired, then the weight of its input links can be increased, making it easier for the unit subsequently to reach its threshold value.

In practice, a PDP model ‘learns’ in this way not through modifying just the weights of the links feeding directly into the output layer, but by adjusting all of the links that indirectly affect the output layer. Modifying the behaviour of a unit in the output layer can therefore involve modifying the weights of links between the hidden and output layers, and between the input and hidden layers. It is possible to adjust the weights of the network in this way automatically, by calculating backwards from the links between hidden and output layers to the links between input and hidden layers. This process of adjusting weights backward from the output layer is known as the **backward propagation of error**.

In addition to their ability to learn, PDP models have other characteristics that are important for our understanding of cognition:

- 1 PDP models need not contain any explicit rules, but can still behave as if they are following such rules. For example, Prince and Smolensky (1997) developed a PDP model of how words are arranged in order to produce grammatical sentences. The behaviour of the model could be explained in terms of a set of grammatical rules, but none of these rules were explicitly present in the model. The behaviour of the model *conformed* to, but was not *governed* by, any explicit rules.
- 2 If a network is damaged, for example through the removal of specific units, performance of the model will typically deteriorate slightly, rather than fail completely. This gradual deterioration in performance is referred to as **graceful degradation**. It is also a characteristic of the human brain – for

example, the gradual loss of neurons in the brain due to ageing has a gradual rather than sudden effect on cognitive processes such as learning and memory.

- 3 PDP models can exhibit emergent properties. An **emergent property** is a behaviour that the model comes to exhibit over time, through training, that was not explicitly programmed into the model. This feature also makes PDP models interesting from a psychological viewpoint as much of human learning derives from accumulated experience rather than explicit instruction.

## 1.2 Rule-based systems

PDP models can be contrasted with rule-based systems. A **rule-based** system models cognition as an explicit set of rules – for example, production rules – that provides a recipe for how the model should behave. **Production rules** contain two parts – a condition and an action. They are structured in the form ‘IF condition is met THEN perform action’. The **condition** specifies what must be true in order for the rule to be applied. The **action** is what the production rule should perform if the condition is true. If a production rule matches its condition and performs its action, then the rule is said to have **fired**. A rule-based model is constructed out of a set of production rules that together can produce the desired cognitive behaviour. The production rule written in an English-like form in Figure 16.5 could be used in a model of how to make a cup of tea. A complete model of tea making might contain production rules for boiling the kettle, warming the pot, adding tea bags to the pot, adding sugar and stirring the tea. Each production rule would fire when the current state of the tea-making process matched its condition. The rule in Figure 16.5 would only fire once the milk had been added to the cup and the tea was ready to pour.

IF the cup just contains milk **AND** the tea in the pot is ready  
**THEN** pour tea into the cup

**Figure 16.5** A production rule from a model of tea making

It is possible to model complex cognitive processes using production rules. For example, a rule-based model of how humans produce grammatical sentences may have production rules for selecting an appropriate verb depending on the meaning of the sentence, and for selecting an appropriate ending for the verb depending on the tense to be used.

Unlike PDP models, rule-based models focus on how the cognitive tasks humans perform can be understood as the processing of information without considering how such processing might be realized in the brain. The relationship between the approaches taken by PDP and rule-based systems can be understood in terms of Marr’s (1982) levels of explanation (discussed in Chapter 1). Marr argued that psychological explanations can be understood at any of three levels: computational, algorithmic and hardware. The PDP approach places a greater emphasis on the hardware level, arguing that if the model reflects some basic properties of the human brain, interesting psychological behaviour will emerge. On the other hand, a wholly rule-based approach places virtually no emphasis on the actual brain, effectively saying that the way the brain works is more a matter for biology than psychology.

Instead, it argues that psychological phenomena can be most appropriately explained at the computational and algorithmic levels.

### 1.3 Cognitive architectures

Newell (1973) argued that it was not sufficient to develop a collection of discrete models to describe a broad range of psychological phenomena. Instead, there should be some integration and consistency across the models being developed. For example, playing chess, recognizing objects and producing grammatical sentences are all psychological processes that use long-term memory. If rule-based models of these three psychological phenomena contained completely different ways of representing, organizing and retrieving from long-term memory, this would clearly be a problem. Humans use the same cognitive processes across a range of tasks.

This led to the development of rule-based cognitive architectures that could account for a range of cognitive processes using the same modelling components. A **cognitive architecture** is an overarching framework that can account for a number of phenomena using a fixed set of mechanisms. As well as maintaining consistency across a set of models, cognitive architectures have another important advantage. Cognitive architectures distinguish clearly between the cognitive model and the computer (and any associated programming languages) on which the model is running. When a cognitive model is developed using a standard programming language (such as the C programming language) there is a need to distinguish which parts of the program are psychologically relevant and which are just dependent on the programming language being used. For example, a programming language has facilities for storing information but the model developer is not necessarily claiming that the way this information is stored in the computer bears any relation to the way humans store information in memory. A rule-based cognitive architecture is ‘emulated’ on a computer, but there is a clear distinction between the working of the model and the working of the computer. By this, we mean that the cognitive architecture is run on a computer but has its own self-contained set of processes, such as production rules. It does not directly use the general purpose processes of the computer that are used to provide the computer user with all kinds of facilities from email to word processing.

PDP is itself a cognitive architecture comprising a fixed set of artificial neural mechanisms. Two of the most well known rule-based cognitive architectures are ACT-R (Anderson and Lebiere, 1998) and Soar (Newell, 1990). In the next three sections we will look in detail at ACT-R, as a rule-based cognitive architecture and some of the empirical data it has been used to model. Particular consideration will be given to the extent to which ACT-R meets Newell’s (1990) goal to develop a cognitive architecture that can provide an integrated and consistent account of a wide range of psychological processes.

## Summary of Section 1

---

- Cognitive modelling involves building a model of a cognitive process and then comparing the behaviour of the model against human performance.

- The two main types of cognitive modelling are parallel-distributed processing (PDP) and rule-based systems.
- PDP models have neural-like properties and aim to demonstrate how a neural architecture can support human cognition.
- Rule-based models focus on how information is processed and give less consideration to how cognitive processes are realized in the brain.
- Rule-based cognitive architectures such as ACT-R and Soar attempt to provide an integrated account of a range of cognitive theories and empirical findings.

## 2 An overview of ACT-R

To give you a taste of cognitive modelling, in this chapter we will describe, evaluate and use Anderson's ACT-R cognitive architecture (Anderson and Lebiere, 1998). ACT-R is perhaps the most widely used cognitive architecture in the cognitive modelling community, and reflects a trend towards **hybrid models** that attempt to span Marr's levels of explanation. Although ACT-R is primarily a rule-based cognitive architecture, it has certain characteristics more usually associated with PDP.

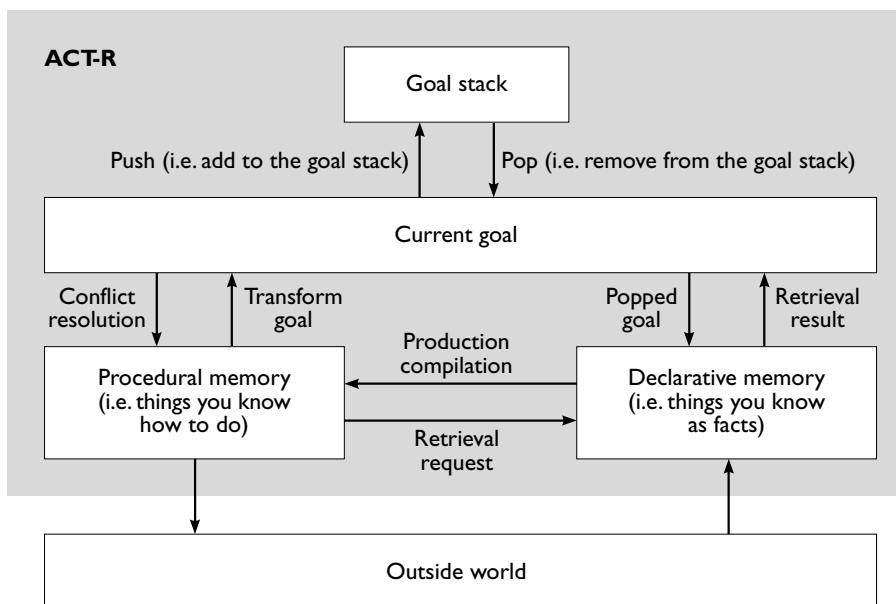
### 2.1 A brief history of ACT-R

ACT (which stands for 'Adaptive Control of Thought') has its roots in Anderson and Bower's (1973) theory of human associative memory, and their model of it, called HAM. A number of different versions of ACT have been developed. The first version of ACT proper, called ACTE, combined elements of HAM's memory representations with rule-based production systems that model control of behaviour and more complex activities like problem solving. In 1983, Anderson revised ACTE producing ACT\* (pronounced 'act star'), which revised the underlying memory system to be more plausible biologically, and introduced a mechanism for learning new rules for the first time. ACT\* was the first complete theory in the ACT series, and was capable of modelling a wide range of behaviours, from memory to complex problem solving and skill acquisition.

In 1993, ACT-R was developed and since then it has been gradually revised. The 'R' stands for 'rational', and refers to Anderson's (1990) theory of 'rational analysis' (recall the different explications of rationality discussed in Chapter 12). Basically, **rational analysis theory** states that each component of the cognitive system is optimized with respect to demands from the environment, given its computational limitations. ACT-R also shifted the emphasis towards a finer-grained model. Whereas earlier rule-based models tended to use a small number of complex rules, in ACT-R there is a definite shift to a larger number of simple rules. Anderson and Lebiere (1998) in fact argue that the simple chunks of the ACT-R's memory, and the simple rules of its production system, are 'atomic' in the sense that they should not be broken down into further, more fine-grained ACT-R constructs.

## 2.2 The architecture of ACT-R

The ACT-R cognitive architecture comprises a clear set of components, whose interactions lead to its special behaviour. These components are more or less distinct modules within it. We have shown an overview of the architecture for ACT-R in Figure 16.6 below.



**Figure 16.6** An overview of the ACT-R architecture

Source: adapted from Anderson and Lebiere, 1998

The ACT-R architecture makes a strong distinction between two different kinds of memory, **declarative memory** and procedural memory (see Chapter 8). Put simply, declarative memory is full of things that we just know, for example, that '1 + 3' is '4', that grass is green and that books have pages in. Procedural memory, on the other hand, is full of things that we know how to do. For example, few people would know the answer to '154 + 367' directly, but most would know a procedure that can be used to work out the answer. The steps in this procedure would be stored in procedural memory.

Importantly, declarative memory and procedural memory are not independent, but interrelated. There are two kinds of connection between them. First, a production rule fired in procedural memory may require elements from declarative memory. This route is shown in Figure 16.6 as a 'retrieval request'. Second, new production rules can be created in procedural memory from 'chunks' in declarative memory (this process will be discussed in Section 2.4). This process is known as **production compilation**. The importance of production compilation can be seen in ACT-R models of learning, where new rules are formed as learners become more skilled at solving problems. For example, an intermediate chess player will pick up new rules through experience, not of the rules of the game itself which they

already know, but rules about how to defend against or exploit particular situations on the board.

There is more to the ACT-R architecture than the interplay between declarative and procedural memory, though. There is also a **goal stack** and a **current goal** (see Section 2.5). These are important for models involving lots of rules, as the current goal is a kind of focus of current attention – this represents what ACT-R is currently trying to do. The stack contains other goals; these are goals that are not the immediate focus of attention, but that still need to be dealt with some time later.

As described so far, ACT-R is isolated from the external world. However, it is worth knowing that attempts have been made to encompass and extend ACT-R even further, providing it with a perceptual and motor system, including motor, speech, vision and audition modules. In fact, Salvucci (2001) has used ACT-R to study the effects of using a mobile phone while driving.

This gives you the big picture of ACT-R. Now let's go into each part of the architecture in a little more detail.

### 2.3 Declarative memory

All of the ACT models, going way back to HAM in 1973, conceived of declarative memory as a collection of **chunks**, or declarative memory elements, which themselves contain a number of elements, usually between two and four. A typical chunk is shown in Figure 16.7.

Chunk1	isa ADDITION-FACT Number1 = 1 Number2 = 3 Sum = 4
--------	--

**Figure 16.7** An example ACT-R chunk

This chunk encodes the addition fact, ‘1 + 3 is 4’. Each of the main rows in the chunk holds a value in a different **slot**. Chunks also have the ‘isa’ slot (pronounced ‘is a’), which says what kind of chunk it is (i.e. ‘this chunk isa addition fact’).

Chunks are not isolated from each other; they are all linked to each other through their values in a kind of network. For example, the chunk shown in Figure 16.7 will be linked to all other chunks that are addition facts or that use the numbers one, three or four. Chunks influence one another through these links by a limited form of spreading activation (discussed in detail in Section 3.3). Put simply, each chunk has a level of **activation**, a kind of energy attached to it. But activation in one chunk tends to leak out and add to all the other chunks it is connected to through its values. For example, increasing the activation of the addition fact ‘1 + 3 = 4’ would automatically increase the activation of other addition facts and of chunks which contain ‘1’, ‘3’, or ‘4’ in their slots. Chunks that are both addition facts and that contain ‘1’, ‘3’ or ‘4’ would get even more activation.

Activation is central to many aspects of ACT-R. If a chunk has a lot of activation, it will be easy to find and to retrieve quickly from memory. If a chunk has very low activation, it will be hard to find, and may never be retrieved at

all. Highly active chunks, therefore, look a bit like short-term memory, and chunks with less activation look a bit like long-term memory. If a chunk has no activation, it has effectively been forgotten, at least temporarily, until its activation level is increased.

Activation is not constant for a chunk. If nothing happens, a chunk slowly loses activation. But if a chunk is used, it gains activation, and the more it is used, the more it gains. This allows ACT-R to model effects such as priming (see Chapter 6). For example, if a chunk is retrieved, its activation will increase. If the chunk is then retrieved again shortly after (before the added activation has been lost) it will be retrieved more quickly. So declarative memory in ACT-R is a lot more than a place for storing chunks. It plays an active and essential role in the behaviour of ACT-R itself.

## 2.4 Procedural memory

The next main component of ACT-R is its procedural memory. This stores procedures in the form of **production rules**, which have a condition (i.e. IF) part, and an action (i.e. THEN) part. Unfortunately, since ACT-R is a computer program, its rules tend to be written in a fairly cryptic form. Figure 16.8 shows a rule, first in an English-like representation, and then in ACT-R form.

Rule in English form	
How to add two numbers together	
<b>IF</b> the goal is to find the answer to <i>number1</i> + <i>number2</i> in a column and we know that <i>number1</i> + <i>number2</i> is <i>sum</i>	
<b>THEN</b> the answer is <i>sum</i>	
Rule in ACT-R form	Commentary
Add-Numbers	The production rule called Add-numbers fires if...
=goal>	the current goal is...
isa ADD-COLUMN	to add up a column of an addition sum in which ...
first-number =number1	there is a first number (=number1) to be added...
second-number =number2	to a second number (=number2)...
answer nil	and the answer is unknown (nil means it is empty).
=fact>	And also we have a chunk in declarative memory...
isa ADDITION-FACT	and it is an addition fact stating that...
addend1 =number1	the first number (=number1) added to...
addend2 =number2	the second number (=number2) gives...
answer =sum	an answer (=sum).
==>	Then
=goal>	In our current goal we can use...
answer =sum	the answer from the addition fact (=sum) as the answer for the goal

**Figure 16.8** An ACT-R production rule in English and in ACT-R form

To give you an idea of the relation between these representations, the condition part of the rule (the IF part) can be found in the rows before the ' $\Rightarrow$ ' character, and the action part (the THEN part) appears after it. The bits that begin ' $=$ something $>$ ' are references to chunks, and each line that follows is a slot name followed by a value. When a word begins with the '=' character, it is a **variable**, that is, it can change each time the rule is used. Variables are very important to production rules, as they are what make rules sufficiently general to run with different problems. For example, the rule in Figure 16.8 could be used to add any addition column for which the matching addition fact was stored in declarative memory. Variables and values will be covered in more detail later.

ACT-R uses production rules in the following way. All specified production rules are available to be used depending on the current state. For a production rule to be used, its condition part (the IF part) must match the current goal and use chunks already available in declarative memory. This gives quite subtle control over timing. For example, the greater the activation of a chunk in declarative memory, the faster productions that use that chunk can be fired. ACT-R uses this technique extensively to give reasonably accurate models of human response times (as we shall see in Section 3).

Activation also plays a significant role in procedural memory. Rules have activation too, and if a rule has a very low activation it might be unused even though its conditions match the current state. However, the more a rule is activated, the more likely it is to be retrieved and used again in the future.

Procedural memory is not fixed like a computer program – new rules can be learned through a process called production compilation. **Production compilation** is the name given to the process by which knowledge is transferred from declarative to procedural memory. The process by which production rules are learned in ACT-R has three stages:

- 1 **Understanding.** The first stage involves understanding instructions and available worked examples. For example, a teacher may provide a child with verbal instructions for multiplying two numbers together. In ACT-R this new knowledge is encoded as chunks in declarative memory.
- 2 **Production compilation.** In the second stage, we try to solve problems by applying these instructions. By working on a range of problems we start to generalize from our experience. For example, through attempting multiplication problems for themselves a child will start to realize the range of problems to which the multiplication instructions can be applied. In the initial stages, the child may only be able to correctly solve the multiplication problem when it has certain characteristics, such as containing or not containing particular numbers. Through experience the child will come to realize how general instructions can be used to solve a wide range of multiplication problems. In ACT-R this is supported by the process of production compilation. Through application, the declarative representation of the instructions becomes transformed into a production rule. The production rule becomes more general than the declarative chunks as specific values in the chunks become replaced by variables in the production rule. For example, chunks representing the instructions for multiplying 27 by 3, may be

generalized into a production rule for multiplying any two digit number by three, and then further generalized for multiplying any two digit number by a one digit number. This process is described in detail in Section 4.1.

- 3 **Practice.** Through practice we solve problems with increasing speed and accuracy. For example, a child having become more experienced with multiplication problems will deliberate less over the individual steps in the task, and come to solve the problems with relative ease. ACT-R explains this in terms of an increasing use of the production rules and a decreasing use of the declarative instructions. When first developed, the production rules may have a low level of activation, making the retrieval and use of the production rules more difficult. Through use, the activation level of the production rules increases until they take over from the declarative representation of the instructions.

Much of ACT-R's explanatory power comes from the production compilation process, which sets up a continual interplay between the chunks in declarative memory and the rules in procedural memory. Instructions for how to perform a given task start out as chunks in declarative memory, and then as performance improves through practice, these chunks are turned (or compiled) into rules in procedural memory, which do the same thing as the chunks but do it automatically.

To complete the circle, as these rules are used, they may themselves create new chunks in declarative memory. As we shall see in Section 4.2, for example, a child who does not know from memory the answer to '4+2' may use a counting procedure to arrive at the answer. The child may then remember this answer and so subsequently can provide the answer to '4+2' directly from memory. In ACT-R this is modelled by a production rule (e.g. the child's counting procedure) producing a chunk in declarative memory (e.g. a chunk representing '4+2=6') that can later be used to answer the same question without using the original production rule. We will come back to ACT-R's approach to learning in Section 4, where we will look at a model that shows it working in practice.

The idea that memory is divided into declarative and procedural memory is central to the ACT-R theory. One of the sources of evidence for this distinction is the experiment conducted by Rabinowitz and Goldberg (1995), which relies on the fact that declarative encodings of instructions can be reversed more easily than procedural encodings of the same instructions (see Box 16.1).

## 2.5 Goals and the goal stack

Production systems like ACT-R's procedural memory are rooted in work on problem solving in the field of artificial intelligence. These systems adopt a goal-directed approach to problem solving. Basically, they take a current goal and a current state, and the system acts either to achieve the goal or add a new goal that needs to be completed first. For example, if someone wishes to have home-made lasagne for their meal, they will probably first set themselves the goal of assembling all the ingredients in their kitchen. This may involve going shopping. Once the initial goal of assembling the ingredients has been met, then he or she can move on to the next stage and make the lasagne using the ingredients.

**16.1****Research study****Rabinowitz and Goldberg's (1995) experiment**

Rabinowitz and Goldberg's (1995) experiment investigated differences between declarative and procedural encodings of the same instructions. If a declaratively encoded instruction can be reversed more easily than a procedurally encoded one, this lends support to the idea of relatively distinct memory systems.

Participants were given a simple alphabet arithmetic task, with stimuli like ' $C + 3 = ?$ ', where the number indicated how many letters the participant should advance along the alphabet. The expected response here would be 'F'. Two participant groups took part in the experiment. The first group was given 12 letter-number combinations in rehearsal, but the second group was given 72 (thereby gaining more practice time). After this practice time, both groups were tested on additive alphabet arithmetic, but also on a transfer task – subtractive arithmetic on the same problems, for example ' $F - 3 = ?$ ', with the expected response 'C'.

Participants who received less practice performed better on the transfer task when the problems presented featured the same number and letters as problems they had tackled in the first task. For example, they answered more quickly to ' $F - 3 = ?$ ' if they had already seen ' $C+3=F$ '. This was not the case for participants that had received more practice.

Participants who received less practice could solve subtraction problems by reversing the addition solutions held in declarative memory. Participants receiving more practice had built a procedure for arithmetic addition within their procedural memory, and this could not be reversed.

In ACT-R, while the current goal (e.g. finding the ingredients) is being undertaken, future goals (e.g. cooking the lasagne) are stored on a **goal stack**. The computer science concept of a stack, on which the ACT-R goal stack is based, has been used commonly for many years. A stack is simply a bit of memory where you can put things in and get them out again, but you can only take them out in reverse order (i.e. last in – first out). Computing also has its own terminology for adding and removing items from a stack. Items are said to be ‘pushed’ onto a stack and ‘popped’ from a stack. This terminology is adopted by ACT-R. Computer stacks can be used to hold a very large number of elements and recall them perfectly. This is also true of the ACT-R goal stack that can store and perfectly recall an arbitrary number of goals. Humans, however, clearly cannot do this – in fact, forgetting a subgoal is a very common ‘slip’ people make in real life.

Because of this, the ACT-R goal stack can be criticized for its lack of psychological plausibility. Anderson and Lebiere (1998) accept this charge and suggest that this is one area in which future work is needed to refine the architecture. Although the goal stack has these problems, Anderson and Lebiere need to think carefully about how to replace it, as it does have the advantage of making the architecture function in a controlled and serial way as each goal is tried one at a time.

## Summary of Section 2

---

- ACT-R is a cognitive architecture with three main components: declarative memory, procedural memory and a goal stack.
- Declarative memory is organized as a set of interconnected chunks forming a network, each chunk having a level of activation.
- Procedural memory is comprised of production rules that also have activation levels and that can perform some action if the condition part of the rule is met.
- ACT-R production memory is not fixed, new rules can be learned, modelling skill acquisition through instruction, examples and practice.
- ACT-R rules depend on a current goal and a stack of pending goals, which make a (questionable) assumption of perfect memory, but which do make high-level cognition serial.

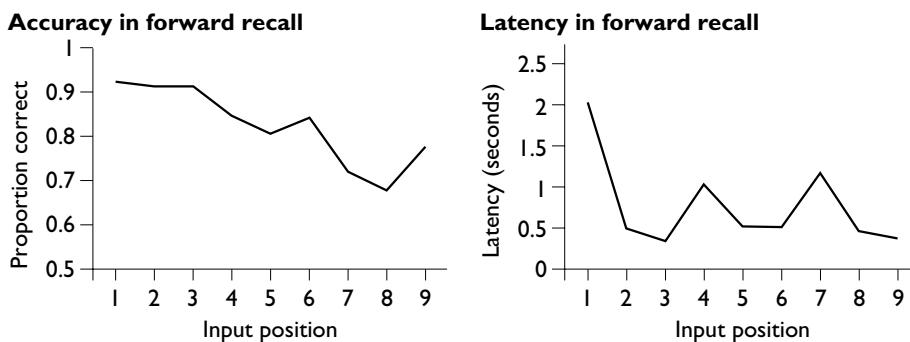
## 3 ACT-R accounts of memory phenomena

As a demonstration of the ability of ACT-R to model human memory, Anderson *et al.* (1998) developed a single model of human performance on list memory tasks.

**List memory** is an experimental paradigm used in cognitive psychology to investigate how people store and recall items from short-term memory. Typically, participants in an experiment are presented visually with a list of items (such as words or numbers) one after another. They are then asked to recall the presented items, possibly after some delay. A restriction may be placed on the order in which the items should be recalled. The participants may be requested to recall the item in the precise order in which they were presented (termed **forward recall**), the reverse order in which they were presented (termed **backward recall**) or in any order (**free recall**).

The ACT-R model of list memory nicely illustrates many key features of the ACT-R cognitive architecture, and list memory has been an active area of ACT-R research in recent years. Although the list memory task is highly artificial, the precise nature of the task and the wealth of empirical data and theoretical explanations of results (e.g. Baddeley, 1986) provide a lot of information to support the building and evaluation of ACT-R models.

Here we will focus on how ACT-R models human performance of forward recall. The human criteria against which the model will be compared are recall latency (i.e. time taken to recall an item) and recall accuracy. Accuracy and latency data in recall for a nine-element list are shown in Figure 16.9. In the empirical study (Anderson *et al.*, 1998) from which this data was collected, participants were initially presented with a string of empty boxes, one to contain each item in the list. Participants were therefore aware of the list length from the beginning of the study. The items were then presented in their appropriate box, one at a time. As one item appeared, the previous item disappeared, so that only one item was visible at any one time. As the last item disappeared, subjects were instructed either to recall the items in a forward



**Figure 16.9** Recall accuracy and latency in the forward recall of a nine element list

Source: Anderson and Lebiere, 1998

or backward direction. Participants would then enter the items into the boxes in order, and could not return to an earlier box once it had been visited.

If we first look at the graph for recall accuracy, two important features should be noted. First, recall accuracy is highest for the first elements in the list. This is called the **primacy effect** and can be explained by assuming that participants rehearse the first elements of the list during the presentation of the later items. Second, accuracy is higher for the last element than the two preceding elements. This is called the **recency effect** and is thought to be due to the last item still being accessible from memory during the recall phase, even though it has not been rehearsed, as its activation level has not decayed. Turning to the graph of recall latency, it should be noted that recall is slower for elements one, four and seven. This is conjectured to be due to the way the items are chunked in declarative memory.

In order to accurately reflect the empirical data, the ACT-R model needs to:

- (a) have a representation of how items are chunked in declarative memory,
- (b) have production rules for the rehearsal of items and retrieval from memory, and
- (c) model activation levels and show how they affect recall accuracy and latency.

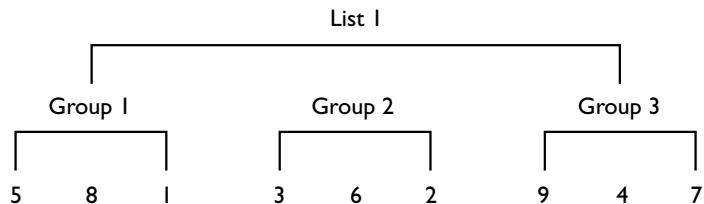
These three points will be considered in the following three subsections.

### 3.1 Declarative representation of lists

Within the ACT-R model of list memory, the list itself is represented in declarative memory as chunks. Chunks are used to represent a list as a set of groups. A **group** can contain as little as two and as many as five or six items.

The way that people mentally group telephone numbers could be represented as ACT-R chunks. Consider for example the main switchboard number for The Open University in Milton Keynes. Written without any spaces to indicate groups the number is 01908274066. Some people, particularly those familiar with the Milton Keynes dialling code, will group the first five items (01908). Individual differences are found in how people tend to group a six digit telephone number, either as two sets of three, or three sets of two. This gives two common groupings of the number, either into three (01908 274 066) or into four (01908 27 40 66) groups.

In a list recall task, participants often organize the list into three groups. This is found to optimize the number of items that can be remembered. For example, the list 581362947 is often grouped as shown in Figure 16.10. To distinguish it from any other lists, we shall refer to it as List1. The three groups are referred to as Group1, 2 and 3. The model of Hitch *et al.* (1996) also represents a list as a set of groups each containing a small number of items.



**Figure 16.10** Organization of a nine-element list into three groups of three

When grouping lists in declarative memory, each group and each individual item within the group is associated with a chunk. List1, as grouped above, would be encoded using 12 chunks, one chunk for each of the three groups and one chunk for each of the nine elements. A chunk can therefore hold a group of items or just a single item. (Recall in Chapter 10 that Chase and Simon (1973) found that expert chess players could remember the positions of more chess pieces after a glance than novices. This suggests that, although experts and novices can form the same number of chunks in short-term memory, experts can represent the positions of a greater number of pieces in one chunk.)

## ACTIVITY 16.2

---

How many chunks in total are required to represent a list of 15 elements with

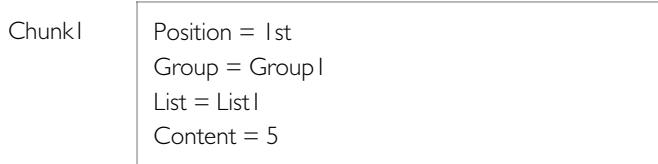
- (a) a group size of three and
- (b) a group size of five?

### COMMENT

- (a) Twenty chunks would be required. Five chunks are required for the five groups of three elements. Fifteen chunks are required for the individual elements. (Note that the list itself does not have a chunk. 'List1' appears as a value in each of the 15 chunks.)
  - (b) Eighteen chunks would be required. Three chunks are required for the three groups of five elements. Fifteen chunks are required for the individual elements.
- 

As mentioned in the previous section, each chunk is represented using **slots** and **values**, however the chunks used to encode groups and individual items have a slightly different set of slots. A chunk associated with individual elements has slots for the group to which it belongs, its position within the group, the overall list to which it belongs and its content (i.e. the list item). The chunk associated with the first item in the list is shown in Figure 16.11. This chunk states that the item in

the first position of Group1 of List1 is the value ‘5’. It is important to note, according to the assumptions of ACT-R, that each individual element has a slot associating the element directly to the list as well as to the group to which it belongs. This allows ACT-R to model some complex memory effects, as we shall see later in Section 3.3.



**Figure 16.11** The chunk associated with the first item of List1

A chunk encoding a group has slots to indicate the list to which it belongs, the number of elements in the group and the position of the group within the list. Each chunk also has a unique name by which it can be referenced. Figure 16.12 represents the chunk associated with the first group of List1. (This group has been given the name Chunk10. While the chunks associated with the nine elements of List1 have been named as Chunks 1 to 9, the chunks associated with the three groups have been named as Chunks 10 to 12.)



**Figure 16.12** The chunk associated with the first group of List1

### ACTIVITY 16.3

---

- Write down the slots and values of the chunk associated with Group3 of List1.
- Write down the slots and values of the chunk associated with the last item of Group2 of List1.

### COMMENT

- According to our numbering scheme, the chunk associated with the third group is Chunk12 and it has a size of three. It also maintains a link to the list. The chunk would therefore look like Figure 16.13:



**Figure 16.13**

- (b) The last element of Group2 would be Chunk6. As this chunk refers to an individual item, it has a slot for its content. The sixth number in the list is 2. The chunk would therefore look like Figure 16.14:

Chunk16

Position = 3rd
Group = Group2
List = List1
Content = 2

**Figure 16.14**

Chunks therefore organize a seemingly flat list of elements into a **hierarchy**, where the list is initially broken down into three groups and the groups broken down into individual elements.

### 3.2 Production rules for the rehearsal and retrieval of lists

The ACT-R model has to perform a number of procedures in order to simulate performance in the list memory task. The list contents have to be rehearsed in memory and then retrieved, and responses have to be given. These procedures are stored in the form of production rules in procedural memory, as described earlier.

The condition part of the production rule may specify what must be the current goal, and may specify what chunks have to be made available in declarative memory. For example, as was shown in Figure 16.8, the Add-Numbers production rule could only fire if the ADD-COLUMN goal was the current goal and the appropriate ADDITION-FACT chunk was in declarative memory. The action part of the production rule either transforms goals held in memory or performs an action to the world outside the model, such as typing a retrieved word onto a computer screen.

**Goal transformations** can be of three types. A goal can be modified, created (i.e. ‘pushed’ on, or placed on top of, the goal stack) or removed (i.e. ‘popped’ from, or removed from the top of, the goal stack). The ACT-R model of list memory comprises a number of production rules for rehearsing, retrieving and giving responses to the questions or tasks of the experiment. For example, Figure 16.15 represents the English form of the production rule for getting the next item from a group (the ACT-R textual syntax has been removed for clarity). Basically, the rule states that if you can retrieve the item you are trying to recall (i.e. the item at position X in group Y), then set a subgoal to output (i.e. say) the retrieved item and then move on to the next item (i.e. item X+1). The action part of the rule therefore modifies the current goal stack by adding a new goal (output item X) and modifying a goal (now look for item X+1, rather than X).

Rule in English form

Get the next item from Group

IF the goal is to get element at position  $X$  from group  $Y$   
and we can retrieve the element  $Z$  at position  $X$  from group  $Y$   
THEN create a subgoal to output the item  
and modify the current goal to look for item  $X+1$

**Figure 16.15** Production for getting the next item from a group

Figure 16.16 shows the production rule used to output an item in a recall task. Here the action part just provides an output and does not transform any goals.

Rule in English form

Type out the item

IF the goal is to output item  $X$   
and we can retrieve a key for  $X$  that is found on the keyboard  
THEN output by pressing the key

**Figure 16.16** Production for outputting an item via a key press

The production rules shown above contain certain characters written in italics ( $X$ ,  $Y$  and  $Z$ ). These are variables that act as empty slots and can accept a range of specific values. These are used in order to satisfy the ACT-R assumption that production rules should provide a level of **generalization** and be applicable to a range of specific cases. For example, the ‘Get next item from Group’ production rule can be used to get any item from a group. Similarly the ‘Output the item’ production rule can be used to output any printable character. Goals will be discussed in further depth in Section 4 on arithmetic skills.

### 3.3 List activation

As we discussed in the previous section, retrieval from declarative memory is performed within the condition part of the production rule. The success of this retrieval process for recall tasks is a function of the activation level of the chunk that matches the condition. The higher the activation level of the chunk the more easily – and faster – it can be retrieved.

Activation is part of the PDP-like nature of the ACT-R architecture. Activation of (artificial neuronal) elements is used within PDP architectures and is itself inspired by our understanding of neurology. The activation level of each chunk in declarative memory is calculated using a set of **activation equations** provided and modifiable within ACT-R.

The activation level of a chunk is calculated as the sum of its base-level activation and associative activation. The **base-level activation** of a chunk depends on the number of times it has been rehearsed in memory and the amount of time that has elapsed since it was last rehearsed. If a chunk has been rehearsed a high number of times and only a small amount of time has elapsed since the last rehearsal, then its base-level activation will be high.

The base-level activation provided by the ACT-R architecture can help to account for primacy and recency effects in list memory. The primacy effect is due to the number of times the item has been rehearsed, increasing its base-level activation. The recency effect is due to the small time lapse since the presentation or last rehearsal of the item, which also means base-level activation will be high.

The **association strength** is the strength of the bond between an item and the required chunk, and influences the flow of activation between chunks. Looking back to Figure 16.10, there will be an association strength between item ‘5’ on the list (the first number on the list) and a chunk that encodes that item, such as Chunk1. The strength of association between an item and a chunk depends on the total number of associations that the item has. An important assumption of the ACT-R architecture is that activation is a limited resource. If an item is only associated with one chunk, then this chunk receives the full associative strength of the item and, therefore, the full effect of any activation. If the item is associated with three chunks, then the association strength is split three ways and less activation will flow to each individual chunk.

The limited capacity of association strength can be used to explain what is known as the fan effect (Anderson, 1974). The **fan effect** is the empirically observed finding that the greater the number of facts related to some concept that a subject has to memorize, the slower the subject will be to recall any one of them. For example, imagine you are asked to memorize three facts about a pretend person called Fred – that he is six feet tall, has a beard and works in a hospital. Your recall will be slower for the fact the Fred is six feet tall because you have been asked to remember other facts about him. If you had been asked to remember just this one fact about Fred, your recall would be quicker. Although the fan effect can be explained in terms of spreading activation, the mechanism within ACT-R is precise and restricted – activation only spreads to the immediate neighbours in the network. However, spreading activation is assumed to encompass a wider region of neurons (or units), than just the immediate neighbours of the activation source.

The fan effect also applies to list memory. In Section 3.1 it was shown that each chunk encoding either a group or an individual item in a list has an association to the list itself, in that case, List1. The association strength for List1 has to be shared out among all the associated chunks. The more chunks that are associated with List1 (either due to a smaller group size or larger list size) the more thinly the association activation has to be spread between the chunks, making it increasingly difficult to (quickly) retrieve any of the associated chunks. Limited association strength therefore offers an account of how list size affects the recall of items from a list. The larger the list the smaller the percentage of items successfully recalled. This is also one of the reasons why the group size is optimally set to three rather than two as this reduces the number of associations with the list concept itself, without overloading any particular group.

In order for a chunk to be retrieved at all, its activation (which is the sum of the baseline activation and the association activation) has to reach a certain level, specified in the ACT-R model. This pre-set level is called the **activation threshold**. A chunk that falls below the activation threshold is unavailable for retrieval by the production rules. The activation level therefore affects recall success. An equation in

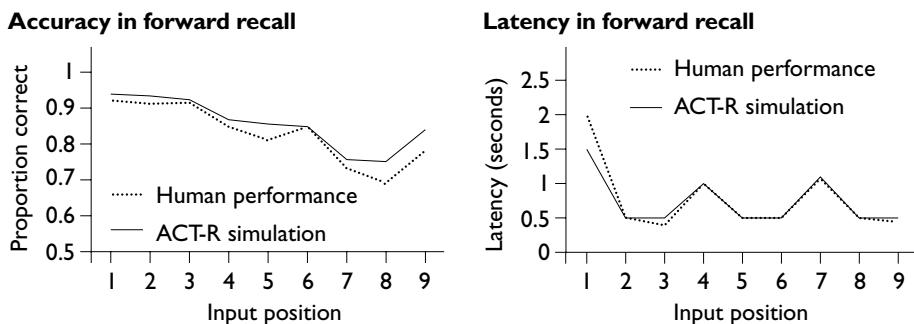
ACT-R also specifies the relationship between activation and latency. The weaker the activation, the slower the recall process will be.

However, retrieval is not always ‘all or nothing’. It is possible to select a chunk that only partially matches the item sought in the condition part of the production rule. This process called **partial matching** happens if, for example, the partially matching chunk has a high level of activation and a fully matching chunk is either absent or below the activation threshold. Partial matching is also important in modelling certain empirically observed effects. **Positional confusions**, where the participant recalls a correct item but in the wrong position, are common in list recall data. (Positional confusions are also discussed in Section 4.1 of Chapter 9, where they are referred to as transposition errors.) For example, in the case of List1, recalling the number 4 in the seventh position rather than the eighth. Once again, equations in ACT-R determine the likelihood of positional confusion. Items are more likely to be confused if they appear in the same group, and if they appear in adjacent positions in the same group. ACT-R uses this mechanism to successfully model the positional confusions exhibited by human participants.

### 3.4 Running the model

We have just covered three important features of ACT-R. First, we have seen how lists can be represented in declarative memory. Second, we have seen how production rules can be used to retrieve items from memory. Third, we have seen how ACT-R employs a model of activation that influences recall accuracy and latency. Now we will consider the running of the ACT-R model of list memory to see how these features work. Figure 16.17 shows the performance of the ACT-R model on forward recall superimposed on the empirical data presented in Figure 16.9.

The results for recall accuracy from the empirical study and the running of the ACT-R model are very similar (Figure 16.17, left). Both show a primacy effect, having the highest recall for the first three elements. Both also show a recency effect for the last item in the nine-element list. The simulation closely mirrors the findings but certain parameters had to be set within the ACT-R model in order to fit



**Figure 16.17** A comparison of ACT-R simulation and human performance on accuracy and latency in forward recall

Source: Anderson and Lebiere, 1998

the results so closely. This included the parameter that affects the likelihood of a positional confusion. The issue of setting parameters in ACT-R and its implications for the validity of the model are considered in Section 3.5.

When examining latency (Figure 16.17, right), data from the empirical study show spikes at intervals of three items that are mirrored in the ACT-R model. In ACT-R, when a production rule is retrieving the next item from a list it takes longer if that item is from the next group rather than from the same group. If the next item is from the same group then the production rule just needs to increment the counter in the current group and retrieve the item. If the next item is from the next group then production rules have to be used to retrieve the next group and then retrieve the first item of that group. So, these data support the assumption that items in declarative memory are grouped, and that groupings of three tend to be used for list recall experiments.

### 3.5 Evaluation of the ACT-R approach to modelling memory

ACT-R has been used to develop versions of a model that can explain a range of empirical data related to list memory. The versions however, are not identical in every way. The overall model described above has been used to model data from forward recall, backward recall and also free recall, but in each case the model needs to be customized in certain ways to fit the data.

Certain parameters need to be set, particularly those relating to the activation settings of ACT-R within the model. For example, parameters affecting the likelihood that an item will be recalled and the time delay involved in accessing a new chunk vary across the different versions of the model. Anderson *et al.* (1998) admit this variation and offer explanations, but it is clearly an area of concern requiring further work. If such variation is required to explain different list memory experiments, which use a heavily restricted and artificial task, to what extent can ACT-R hope to provide a unified theory of cognition?

However, ACT-R does impose certain architectural constraints, limiting how the model can work. These include the procedural–declarative distinction and the use of chunks to group items in declarative memory. However, these constraints still allow for some flexibility when modelling empirical data. Decisions on how to deal with this flexibility are called auxiliary assumptions. An **auxiliary assumption** is made on a case by case basis to deal with the peculiarities of a particular experiment. How the rehearsal of previously presented items in a list occurs and how this competes with attention to the presentation of the remaining items in the list is an example of an auxiliary assumption. These can be contrasted with architectural assumptions. An **architectural assumption** (such as the procedural–declarative distinction) makes a general claim as to the nature of human cognition, and is consistent across all cognitive models developed using the architecture. If an architectural assumption of ACT-R made it impossible to model certain empirical data, then this would suggest that the assumption does not reflect a general feature of human cognition, and that the architectural assumption should be rejected or at least modified.

There is ongoing debate as to whether ACT-R (and other cognitive architectures) sufficiently constrain the modelling process. However, one advantage of ACT-R and other cognitive architectures is that at least all assumptions are made explicit, allowing such debates to occur.

## Summary of Section 3

---

- Chunks are used to organize items in declarative memory into groups.
- Chunks have slots and values and are used to encode both groups and individual items.
- Chunks have an activation level comprising baseline activation and association strength.
- ACT-R is used to model the accuracy and latency of forward and backward recall, as well as other list memory experiments.
- Model fitting is used to match the ACT-R model to the empirical results.
- Any model makes associated architectural and auxiliary assumptions that jointly specify the model and how it works.

## 4 Learning and using arithmetic skills

In the previous section, we saw ACT-R’s model of list memory, which focused particularly on how declarative memory items are represented and how their retrieval is affected by levels of activation. In this section, we will focus more on production rules in ACT-R and their role in the modelling of problem-solving behaviour and the learning process as a novice acquires expertise through practice.

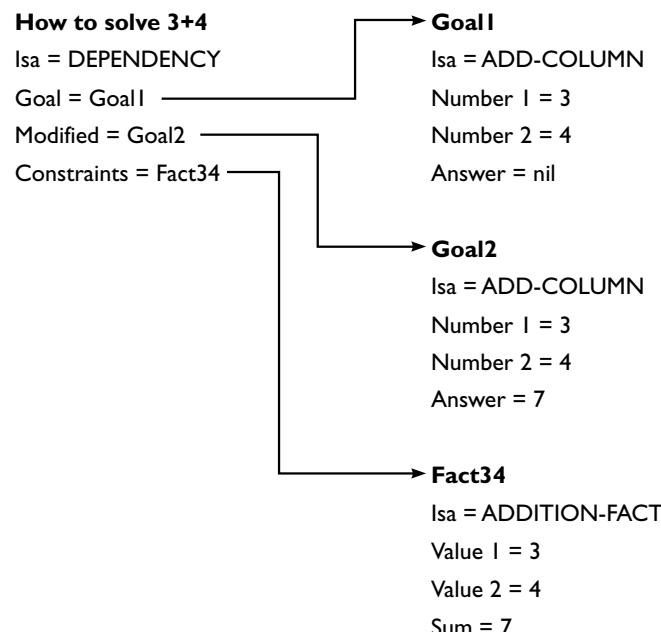
### 4.1 Production compilation

As we discussed in Section 2.4, the ACT-R approach to learning comprises three stages. In the first stage of learning, instructions that the learner has been given are encoded as chunks in declarative memory. A separate chunk is used to represent each step in the instructions. Chunks that represent steps in a process (rather than facts) are called **dependency goal chunks**. A dependency goal chunk is created every time a goal from the goal stack has been successfully completed. For example, if the goal is to find the answer to ‘ $3+4$ ’, and this is solved by matching the goal with the addition fact ‘ $3+4=7$ ’, then a dependency chunk would be created. This dependency chunk would in effect say:

If the goal is to find the answer to  $3+4$  and there is an addition fact  $3+4=7$   
then the answer is 7

This dependency goal chunk is shown to the left of Figure 16.18. The other chunks that it refers to in declarative memory are shown to the right of the figure. Any

dependency chunk represents how an unsolved goal can be turned into a solved goal by using one or more other chunks in memory. The dependency goal chunk in Figure 16.18, which for clarity we have labelled ‘How to solve 3+4’ states that the unsolved ‘Goal1’ ( $3+4=?$ ) was turned into the solved ‘Goal2’ ( $3+4=7$ ) using ‘Fact34’ ( $3+4=7$ ).



**Figure 16.18** An example dependency for adding numbers in arithmetic

Each dependency goal chunk can be thought of as representing a lesson learned from experience, and the lessons represented in dependency goal chunks are very specific. The dependency goal chunk in Figure 16.18 describes how to solve ‘ $3+4$ ’ and no other addition sum. Dependency goals become generalized through production compilation. The first step of production compilation turns a specific dependency goal chunk into a specific production rule. The second step turns the specific rule into a general rule. These two steps will be considered in turn.

First, production compilation works by turning the dependency goal chunk into a new production rule. This produces a new rule that has the unsolved goal and the chunks that were used to solve the goal as the IF part of the rule. The solved goal is placed into the THEN part of the rule. If applied to the goal dependency chunk shown in Figure 16.18, the rule in Figure 16.19 would be produced. For illustration, the production rule has been given the name Add-3-to-4.

So far so good, but this produces a rule that would only work for the problem ‘ $3 + 4$ ’. The second part of product compilation involves generalizing the rule so that it can be used to solve a wider range of goals. ACT-R achieves generalization by replacing specific items in the rule by variables. The heuristic used by ACT-R

Rule in ACT-R form	Commentary
Add-3-to-4	The production rule called Add-3-to-4 fires if...
=goal>	the current goal is...
isa ADD-COLUMN	to add up a column of an addition sum...
first-number 3	and the first number is 3...
bottom-number 4	and the second number is 4...
answer nil	and the answer is unknown (nil means it is empty).
=fact>	And also we have a chunk in declarative memory...
isa ADDITION-FACT	and it is an addition fact stating that...
addend1 3	the number 3 added to...
addend2 4	the number 4 gives...
answer 7	an answer 7.
==>	Then
=goal>	in our current goal we can put...
answer 7	the answer as 7.

**Figure 16.19** A specific production rule produced from the dependency goal chunk shown in Figure 16.18

is that if the same item appears in two or more places within the rule then these should be replaced by a variable. For example, in Figure 16.19, the item ‘3’ appears in two places. It appears as the first number in the goal, and as addend1 in the addition fact. Production compilation would replace these with a variable. This means that the items in first-number and addend1 would have to be the same, but not necessarily ‘3’. A similar transformation would happen to the two occurrences of ‘4’ and the two occurrences of ‘7’. This would in effect produce a rule that said:

If the goal is to find the answer to X+Y and there is an addition fact X+Y=Z  
then the answer is Z

The rule that emerges from this process would work just as well for ‘2 + 2 is 4’, and ‘5 + 2 is 7’, as it does for ‘3 + 4 is 7’. The newly made rule, created by generalizing the rule shown in Figure 16.19 is the one we saw earlier in Figure 16.8. So, ACT-R does not have to be programmed to deal with every eventuality, it can learn new rules to deal with new examples.

Of course, generalizing a rule like this may not always work properly. Although the new rule has become part of procedural memory, its use depends on its activation. If the rule works, it will gain activation and will become an active player in problem-solving actions. On other hand, if it doesn’t work, it will gradually lose activation and become less significant, until it is eventually forgotten.

Overall therefore, production compilation transforms a specific ‘lesson learned’ represented as a dependency goal chunk into a general ‘lesson learned’ represented as a production rule in procedural memory.

## 4.2 An example of human problem-solving behaviour: addition by counting

This gives you a pretty good idea of how a single step in problem solving can be learned, but perhaps this is focusing on too small a part of the problem. Now let's look at how this can be extended to learn a more realistic skill, addition by counting. In looking at this, we are turning the clock back a bit from the examples we have shown before, to the point where we have a child who does not yet know that '3 + 4 is 7'. In ACT-R terms, there is no ADDITION-FACT for '3+4=7' in declarative memory.

This introduces a whole new issue: how do children learn these addition facts? One possibility is that they are simply learned by rote – children recite them often enough that they become chunks in declarative memory directly. However, there is a second possibility: that they are developed through a procedure. Basically, given the problem: 'What do you get if you add 3 and 4?' a child starts from '3' and adds '1' to it four times. The result is, of course, '7'. Siegler (1988) provides an account of how children use this counting strategy when learning arithmetic.

The full ACT-R model for this would be a bit hard to read and to understand, so let's look at a formatted version of it (see Figure 16.20). Above each rule is a plain English description of what it does.

These rules may look clumsy but they are enough to model children working out the answer to problems like '3 + 4 is 7', as if by counting on their fingers. Let's look at how these rules work in a little more detail.

First of all, the rules 'Add-numbers' and 'Subgoal-counting' work together. Basically, if the child already knows the answer (for example, the answer is a chunk in declarative memory, as an addition fact, for example) then the child can just say it (that's what 'Add-numbers' does). On the other hand, if the child does not know the answer, they have to begin to work it out. The three rules 'Start-count', 'Add1-count', and 'Stop-count' all work together, once counting is started by the rule 'Subgoal-counting'. 'Start-count' starts the counting process, then 'Add1-count' is used once for each step and 'Stop-count' stops it when it is complete. Then, the rule 'Found-sum' retrieves the answer from the counting process. Finally, there is a rule 'Say' that is used when the child says the answer to the problem. Figure 16.21 shows how a model might trace through these various rules in practice, given the question 'What is the answer to 3 + 4?'

**Production rules for addition by counting (in English form)**

This is the rule from Figure 16.8 again. It can answer any addition problem for which it can access the appropriate fact from declarative memory:

<b>Add-numbers</b>	<b>IF</b> goal is to answer a question about the sum of $N_1$ and $N_2$ and can retrieve as a fact the sum of $N_1$ and $N_2$ <b>THEN</b> replace current goal with one to say the sum
--------------------	--

If the sum cannot be retrieved because the addition fact is not in declarative memory then this rule will create a subgoal to calculate the answer by counting:

<b>Subgoal-counting</b>	<b>IF</b> goal is to answer a question about the sum of $N_1$ and $N_2$ <b>THEN</b> create subgoal to calculate the sum of $N_1$ and $N_2$ and create subgoal to say the sum
-------------------------	--

If the goal is to calculate the answer by counting then this rule turns the goal 'Add X to Y' into the goal 'Add 1 to X, Y times':

<b>Start-count</b>	<b>IF</b> goal is to calculate the sum of $N_1$ and $N_2$ <b>THEN</b> create subgoal to add 1 to $N_1$ , $N_2$ times
--------------------	---

This can add 1 to the number and create a new goal (for example 'Add 1 to 4, 3 times' can lead to the goal 'Add 1 to 5, 2 times'):

<b>Add1-count</b>	<b>IF</b> goal is to add 1 to $N_1$ , $N_2$ times and New $N_1$ is 1 more than $N_1$ and New $N_2$ is 1 less than $N_2$ <b>THEN</b> change goal to add 1 to New $N_1$ , New $N_2$ times
-------------------	--

This stops the counting process when no more '1's need to be added (for example 'Add 1 to 7, 0 times'):

<b>Stop-count</b>	<b>IF</b> goal to add 1 to a number, zero times <b>THEN</b> mark goal as completed
-------------------	---

This retrieves the answer from the counting process:

<b>Found-sum</b>	<b>IF</b> goal is to find the sum of $N_1$ and $N_2$ and can retrieve the calculated sum of $N_1$ and $N_2$ <b>THEN</b> mark goal as completed
------------------	--

This prints the answer on the computer screen:

<b>Say</b>	<b>IF</b> goal is to say $N$ <b>THEN</b> output $N$ to the screen and mark current goal as completed
------------	--

**Figure 16.20** Production rules for addition by counting

Source: based on Anderson and Lebiere, 1998

<b>ACT-R trace</b>	<b>Explanation</b>
Cycle 0: Subgoal-Counting	Create a subgoal to find the answer to three plus four by counting
Cycle 1: Start-count	Set the goal to add one to three, four times
Cycle 2: Add1-count	Add one to three to make four and change the goal to add one to four, three times
Cycle 3: Add1-count	Add one to four to make five and change the goal to add one to five, two times
Cycle 4: Add1-count	Add one to five to make six and change the goal to add one to six, one time
Cycle 5: Add1-count	Add one to six to make seven and change the goal to add one to seven, zero times
Cycle 6: Stop-count	Goal is add one, zero times, therefore the counting goal is completed
Cycle 7: Found-sum	Calculated answer has been retrieved
Cycle 8: Say	Output the answer to the screen
'Seven'	The outputted answer is seven

**Figure 16.21** ACT-R ‘trace’ for ‘What is the answer to  $3 + 4$ ?’ (first time around)

However, if the same question is put again immediately, you might get a completely different behaviour from ACT-R (as shown in Figure 16.22).

<b>ACT-R trace</b>	<b>Explanation</b>
Cycle 9: Retrieve-sum	Retrieve the answer to three plus four
Cycle 10: Say	Output the answer to the screen
'Seven'	The outputted answer is seven

**Figure 16.22** ACT-R ‘trace’ for ‘What is the answer to  $3 + 4$ ?’ (second time around)

## ACTIVITY 16.4

---

Why do you think ACT-R uses a different process to give the answer second time around?

### COMMENT

The second time ACT-R runs, the chunk that records the answer to ' $3 + 4$  is 7' has already been stored in declarative memory, and has a relatively high activation. This means that second time around, the rule 'Add-numbers' can pick up the answer directly.

---

In practice, this general procedure can work out lots of addition facts, given the ability to count up from one. Over time, this means that chunks corresponding to these addition facts will be stored in declarative memory, and the more they are used (in other, more complex, arithmetic problems, for example) the more they gain activation as chunks in their own right, and the rules may be used less and less. The example shown here, where the answer is retrieved directly after only one trial, is a bit unrealistic – in practice one might want ACT-R to go through the process many times before the new chunk can be retrieved reliably.

Of course, this is not necessarily the only way that children learn these arithmetic facts. Although some children do seem to use this process (Siegler, 1988), others may learn them by rote. Modelling can help to reveal the implications of these strategies in a way that makes it easier to design experiments and use other techniques to study them in more detail. This leads to one of the more interesting features of cognitive models – they correspond more to individuals than to populations, and for this reason we sometimes need to be careful comparing a model's predictions and statistical results from an experiment. In the case of the addition by counting example we have just looked at, we can study differences between individuals in problem solving by looking at the differences between the rules that represent their problem-solving strategies.

### **4.3 Models of learning and problem solving in practice**

In the previous two sections we have seen two complementary parts of the ACT-R approach to learning. In Section 4.1 we considered how the ACT-R production compilation process can be used to create general rules from dependency goal chunks. By contrast, in Section 4.2 we have described how production rules, in this case a model of a counting procedure, can be used to create new facts in the form of chunks in declarative memory. So to what extent does ACT-R provide a general explanation of how humans learn?

Learning is modelled by ACT-R as an incremental process through which new knowledge is acquired declaratively, and performance gradually becomes faster and less error prone though practice. Such an account of learning would seem to accurately reflect how, for example, someone learned to change gears in a manual car. The instructor initially gives verbal instructions (i.e. a declarative account) on how to change gears. Initial attempts by the learner to change gear are slow, deliberate, error prone and rely heavily on the declarative account. Months later the same person will be able to swiftly move through the gears, rarely making errors, and will no longer be using the declarative instructions of how to perform the process. An ACT-R account of this form of learning appears highly satisfactory. Performance gradually becoming faster, less error prone and more automatic can be explained as being due to production compilation.

However, other forms of human learning are harder to explain, as learning may not always be characterized as a gradual increase in performance through experience. For example, children's learning of the past tense verbs in English doesn't quite follow this pattern (Bowerman, 1982). Initially children are very good at forming the past tense correctly. However, over time, and as they learn more

words, their performance deteriorates, and they start tending to use the regular verb rule ‘add –ed’ even when they shouldn’t, saying, for example, ‘breaked’ rather than ‘broke’. Of course, over time, they overcome this problem, and the accuracy goes up again – forming a kind of ‘U-shaped’ pattern.

This pattern can be interpreted as being due to a progression through three ways of performing the task during learning. At first, each problem to be solved (e.g. each past tense verb to be constructed) is dealt with uniquely. The high degree of regularity (e.g. the large number of past tense words that can be created by just adding ‘ed’) is not reflected in the cognitive mechanism used to generate past tense verbs. Later, a single mechanism is used to solve an (overly) wide range of cases. The child now starts to use ‘breaked’ rather than ‘broke’. Finally, exceptions are correctly handled and the child starts to say ‘broke’ once again.

A similar U-shaped pattern of performance over time has been found in childrens’ ability to solve certain mathematical puzzles (Richards and Siegler, 1981), and in students’ radiological diagnoses (Lesgold *et al.*, 1988). Lesgold *et al.* found that students roughly three or four years into a course perform worse than more experienced professionals, but also worse than they did the previous year. In terms of performance, it appears that sometimes a skill has to be acquired, lost and then regained. Human learning therefore may be more complex than ACT-R might suggest.

## Summary of Section 4

---

- Complex behaviour can be modelled using production rules, which contain a condition (IF part) and an action (THEN part).
- ACT-R has a production compilation mechanism for learning new production rules from instructions and examples.
- Learning can also involve the creation of new declarative chunks from production rules.
- Learning and problem-solving behaviour in ACT-R, as driven by the production rules, is influenced by levels of activation.
- ACT-R tends to characterize learning as an incremental process through which performance gradually becomes faster and less error prone. Human learning does not always follow this pattern.

## 5 A comparison of ACT-R and PDP

In the past few sections we have described the ACT-R cognitive architecture. Here we will compare ACT-R as an example of a rule-based cognitive architecture against the PDP cognitive architecture overviewed at the beginning of the chapter.

Children’s learning of past tense verbs in English is a good point for comparison, as it has been modelled using both PDP models and rule-based models like ACT-R.

The central question is: what is happening to cause the ‘U-shaped’ pattern effect? One possibility is that there is an area of memory that functions like a PDP network and that can generalize as it learns from examples. Rumelhart and McClelland proposed a model like this for learning past tenses.

In Rumelhart and McClelland’s model, the changes in accuracy are caused by the child learning more new verbs. When there are only a few words, a network can fit in all the words by simply making strong links between individual input and output units. Each past tense verb is therefore dealt with uniquely. There comes a point when this won’t work – when there are more words than available units – and then the network needs to build a more generalized association. For a short time the model over-generalizes, as a child does. This over-generalization is corrected in the PDP network through training (see Section 1.1), where the network receives feedback as to the correct answer.

Rumelhart and McClelland’s two-layer model was relatively simple – it depended on this forced increase in vocabulary. Plunkett and Marchman (1991) used a three-layer network instead, and found that adding a hidden layer produced a similar ‘U-shaped’ pattern without needing to increase the vocabulary dramatically at one point in the learning process. They still needed a gradual increase in vocabulary to get the right pattern, however (Plunkett and Marchman, 1996).

In contrast, Taatgen and Anderson (2002) set out to model past tense learning using ACT-R, where there are two parts of memory involved: declarative memory and procedural memory. They suggested that children initially learn past tenses as declarative chunks, and then through a process of production compilation (see Section 4.1) form a slightly unreliable production rule to generate past tenses. Over time, they develop a blocking mechanism that stops the rule from being used when there is an exception stored in declarative memory. The dip in performance is caused by the unreliability of the production rule when initially constructed.

Both the PDP and ACT-R models correspond, more or less, to the observed behaviour, but the underlying explanations differ in a few subtle ways. For example, the PDP model depends on feedback. However, there is a problem with this explanation because children aren’t always given feedback, and even when children are corrected, they still tend to use the over-generalized regular verb rule. Conversely, the dip in performance of the ACT-R model is due to the unreliability of the production rule when first formed. This unreliability is responsible for cases of over-generalization, but feedback is not required in order to correct this over-generalization.

The PDP and ACT-R models make different theoretical claims. The PDP model has the assumption that there is one area of memory, and change is driven by change in acquired vocabulary. The ACT-R model has the assumption that there are two areas of memory (declarative and procedural), and change is driven by practice. In principle, these differences can be tested, and experiments used to gather empirical evidence on the matter.

To sum up, the key difference between rule-based architectures, such as ACT-R, and the PDP approach is the nature of the representation used in the model. PDP models, being closer to Marr’s hardware level, are described as having a sub-symbolic representation. **Sub-symbolic models** do not contain any explicit

representations of symbols such as production rules. Instead, they construct a sub-symbolic neural-like representation that can help support and explain a symbolic account of cognition. Conversely, rule-based systems explicitly use **symbolic representations**, such as the declarative chunks and production rules we have seen in this chapter.

In terms of their operation, PDP models, as their full name suggests, work in parallel, with signals being passed simultaneously throughout the network of artificial neurons. Symbolic cognitive architectures tend to be serial in operation. In ACT-R only a single production rule fires at any one time. The parallel processing of the network is mimicked in ACT-R by the activation equations that simultaneously update the activation of all elements in declarative memory.

A final important difference between ACT-R and PDP concerns the kinds of cognitive phenomena that they most successfully explain. Sub-symbolic cognitive architectures such as PDP models, although able to emulate rules, are generally stronger at explaining automatic processes (e.g. face recognition). Symbolic cognitive architectures such as ACT-R are stronger at modelling consciously controlled processes such as problem solving.

As we have seen, ACT-R incorporates some features of the PDP approach. However, to what extent is it possible (or desirable) to build a rule-based system fully integrated with a PDP architecture?

This was attempted by Lebiere and Anderson (1993) in the design of ACT-RN, in which aspects of the ACT-R architecture such as chunks and the goal stack were implemented using a PDP network. Although successful to a degree, many features of ACT-R were difficult to model using PDP. This finding led them to actually remove many of these features from ACT-R, on the grounds that they were not neurologically plausible. For example, ACT-R used to allow a slot in a chunk to have a long list of items as its value. This cannot be done in the current version of ACT-R and lists must be represented as described in Section 3.1. The experiment with ACT-RN also led to PDP-like features being incorporated into ACT-R itself. Partial matching, as introduced in Section 3.3 is one such example. There is therefore reasonable justification to refer to ACT-R as a hybrid architecture that shows both PDP and rule-based characteristics.

It is however still unclear to what extent it is possible to completely integrate symbolic and sub-symbolic architectures. And even if it is possible, it may not always be desirable. The models will necessarily be far more complex, and may inherit the weaknesses rather than the strengths of the symbolic and sub-symbolic approaches. This is one of the reasons why the current version of ACT-R has features motivated by PDP, such as activation, but does not encompass all features of PDP within it.

## Summary of Section 5

---

- Rule-based architectures such as ACT-R provide a symbolic account of human cognition, operate in a largely serial way and are particularly strong at modelling consciously controlled processes such as problem solving.

- PDP architectures provide a sub-symbolic account of human cognition, operate in parallel and are particularly strong at explaining automatic processes such as face recognition.
- Attempts to completely integrate symbolic and sub-symbolic architectures have had limited success, but there appear to be advantages in the translation of certain coarse-grained features from one architecture to the other.

## 6 When is a model a good model?

Modelling has a long and respectable heritage within psychology, with computers being used to model cognitive behaviour even at the birth of cognitive psychology itself, from 1956. Throughout, there has been a continuing question about how models fit in with experimental psychology. One very big question in cognitive modelling is: given a model, how do you know whether it is a good model or not? In this section we consider three criteria against which a model can be judged, followed by a description of the Newell Test, which constitutes an ambitious agenda for cognitive modelling. The three evaluation criteria we wish to consider are:

- The extent to which the behaviour of the model fits human performance.
- The validity of the model from the viewpoint of psychological theory.
- The parsimony of the model – the extent to which unnecessary complication is avoided.

In Section 1, we defined cognitive modelling as building a model of a cognitive process and comparing the behaviour of the model against human performance. If the model behaves in the same way as humans, then the structure of the model, and the way it works may give some insight into how humans perform the task. Clearly, if the behaviour of the model does not mirror human performance, then there is no support for the hypothesis that the internal workings of the model reflect human cognitive processes. And, of course, this failure of a model to fit the data can itself be an important and useful lesson learned.

However, it should not be assumed that the closer the fit to the empirical data, the better the model (Roberts and Pashler, 2000). Although a good model should at least roughly approximate to the data, the most closely fitting model is not necessarily the best. As described by Pitt and Myung (2002) a cognitive model could actually over-fit the data. A model may be so carefully customized to a specific set of empirical data that the generalizability of the model and its components to similar cognitive processes has been jeopardized. The extent to which the model fits the empirical data is therefore insufficient on its own as a measure of quality.

This leads us to our second criterion. The internal structure of the model, by which it produces behaviour, needs to be defensible in terms of the psychological literature. As described in Section 1.3, one motivating factor in the development of cognitive architectures was to logically separate the cognitive model from the workings of the computer. The features of ACT-R available to the modeller, such as procedural and declarative memory, chunks, production rules and production compilation are the mechanisms by which the model produces its behaviour. Each of

these features can be debated and compared against the psychological literature. This clear distinction between the model and its computer implementation has been one of the great successes of work into cognitive architectures.

Our third criterion is parsimony. The law of parsimony or **Ockham's Razor** states that an explanation of a phenomenon should not contain any unnecessary detail. Specifically, and in relation to cognitive modelling, a model should contain only the minimum number of components and so should not contain components that do not impact on the behaviour of the model. Therefore, any component of a model has to provide explanatory significance that justifies the additional complexity that it also brings. Ockham's Razor can also be used to criticize the careful fitting of a model to the empirical data, as this can increase the complexity of the model for little gain.

Despite its increasing maturity, cognitive modelling still lacks a clear method for how models should be evaluated. However, the above three criteria can be used to provide a broad evaluation of any cognitive model, whether it be a rule-based or a PDP model.

Other work in the area of model evaluation has aimed to devise and follow an ambitious set of criteria against which individual cognitive models and the progression of the cognitive modelling field as a whole can be tracked. Anderson and Lebiere (2003) elaborate Newell's (1990) 12 criteria for assessing the quality of a model, which they call the **Newell Test**. These are shown in Box 16.2.

## 16.2

### Constraints on a human cognitive architecture (after Anderson and Lebiere, 2003)

A successful model should:

- 1 Behave as an (almost) arbitrary function of the environment (universality)
- 2 Operate in real time
- 3 Exhibit rational (i.e. effective) adaptive behaviour
- 4 Use vast amounts of knowledge about the environment
- 5 Behave robustly in the face of error, the unexpected and the unknown
- 6 Integrate diverse knowledge
- 7 Use (natural) language
- 8 Exhibit self-awareness and a sense of self
- 9 Learn from its environment
- 10 Acquire capabilities through development
- 11 Arise through evolution
- 12 Be realizable within the brain

Anderson and Lebiere give ACT-R a grade for each point and, based on these criteria, there are some areas where ACT-R is strong. It is pretty good at behaving as a function of the environment, exhibiting rational behaviour, at coping with error, learning, and at modelling real-time behaviour. But there are other areas where ACT-R is much weaker, such as in using natural language, exhibiting self-awareness and being realizable with the brain.

These criteria should however not only be used to highlight the strengths and weaknesses of cognitive architectures such as ACT-R and PDP but also show how researchers working with different architectures can learn from each other. Anderson and Lebiere (2003) claim that the Newell Test could lead PDP researchers to incorporate ideas from ACT-R, similar to the way ACT-R has over recent years incorporated ideas from PDP. These criteria and the cognitive architectures they evaluate can therefore help to provide an overarching account.

As mentioned in Section 1.3, Newell (1973) argued that at that time cognitive psychology was asking lots of small questions, and progressing through small steps, but that the big picture was disjointed because there was little in the way of an overarching framework to glue the work together. He argued that a move to complete theories and models rather than partial ones, to complex composite tasks rather than narrow focused ones, and to models that would cope with many tasks rather than just one or two, would help the science of cognitive psychology to progress more effectively. In many senses, Newell's article laid the foundations for cognitive architectures like ACT-R and Newell's own Soar architecture (Newell, 1990). And the move to using cognitive models in conjunction with empirical studies and the development of cognitive theory, to help connect diverse elements of cognitive psychology, is set to continue.

As a set of points, though, Newell's list of criteria is helpful simply because it is so extensive. It shows just how far there is to travel before cognitive models are capable of explaining cognitive behaviour in an integrated manner. But we should not leave models on this note, as this issue and the list of criteria apply to all cognitive psychological theories not just the kinds of model exemplified by ACT-R. ACT-R may still have a long way to go, but it is one of the best approaches available.

## Summary of Section 6

---

- It is possible to set out many criteria to help judge the usefulness of a model.
- ACT-R performs reasonably well across the board, although it shows weaknesses in the areas of natural language, self-awareness and biological plausibility.
- Models can be useful independently of the quality of their empirical predictions, in that they allow a community of researchers to be brought together to share ideas.

## 7 Conclusions

The advent of computers was central to the foundations of cognitive psychology. Computers provided both a new set of concepts that could be used to understand human behaviour and a new method that could be used to study it. Symbolic cognitive architectures such as ACT-R, however, are offering something more precise than computer metaphors of the human mind. Rather they assume that the working of the mind is essentially the symbolic representation of knowledge (e.g. in the form of chunks) and the use and transformation of these symbolic representations in order to perform tasks (e.g. actions performed by production rules). These cognitive architectures are emulated on a computer but can be thought of as distinct from the workings of the computer itself. The PDP cognitive architecture is also emulated on a computer, but here the assumption is that cognitive functions can be constructed from artificial neural elements having some similar properties to the human brain. The development of ACT-RN and the incorporation of PDP-like properties into the ACT-R architecture highlights a trend towards a hybrid approach to modelling that aims to combine the benefits of symbolic and sub-symbolic approaches.

## Further reading

- Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Anderson, J.R. and Lebiere, C. (2003) ‘The Newell Test for a theory of mind’, *Behavioural and Brain Sciences*, vol.26, pp.587–640.
- Taatgen, N.A. and Anderson, J.R. (2002) ‘Why do children learn to say “broke”? A model of learning the past tense without feedback’, *Cognition*, vol.86, no.2, pp.123–155.

## References

- Anderson, J.R. (1974) ‘Retrieval of prepositional information from long-term memory’, *Cognitive Psychology*, vol.6, pp.451–74.
- Anderson, J.R. (1990) *The Adaptive Character of Thought*, Hillsdale, NJ, Lawrence Erlbaum.
- Anderson, J.R., Bothell, D., Lebiere, C. and Matessa, M. (1998) ‘An integrated theory of list memory’, *Journal of Memory and Language*, vol.38, pp.341–80.
- Anderson, J.R. and Bower, G.H. (1973) *Human Associative Memory*, Washington, Winston and Sons.
- Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Anderson, J.R. and Lebiere, C. (2003) ‘The Newell Test for a theory of mind’, *Behavioural and Brain Sciences*, vol.26, pp.587–640.
- Baddeley, A. (1986) *Working Memory*, Oxford, Clarendon Press.

- Bowerman, M. (1982) 'Reorganizational processes in lexical and syntactic development' in Wanner, E. and Gleitman, L.R. (eds) *Language Acquisition, The State of the Art*, Cambridge, Cambridge University Press.
- Chase, W.G. and Simon, H.A. (1973) 'Perception in chess', *Cognitive Psychology*, vol.4, pp.55–81.
- Hitch, G.J., Burgess, N., Towse, J.N. and Culpin, V. (1996) 'Temporal grouping effects in immediate recall: a working memory analysis', *Quarterly Journal of Experimental Psychology*, vol.69A, pp.116–39.
- Lebiere, C. and Anderson, J.R. (1993) 'A connectionist implementation of the ACT-R production system' in *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pp.635–40, Hillsdale, NJ, Erlbaum.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D. and Wang, Y. (1988) 'Expertise in a complex skill: diagnosing X-ray pictures' in Chi, M.T.H., Glaser, R. and Farr, M.J. (eds) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum.
- Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H.Freeman.
- Newell, A. (1973) 'You can't play 20 questions with Nature and win: projective comments on the papers of this symposium' in Chase, W.G. (ed.) *Visual Information Processing*, pp.283–308, New York, Academic Press.
- Newell, A. (1990) *Unified Theories of Cognition*, Cambridge, MA, Harvard University Press.
- Pitt, M.A. and Myung, I.J. (2002) 'When a good fit can be bad', *Trends in Cognitive Science*, vol.6, pp.421–5.
- Plunkett, K. and Marchman, V. (1991) 'U-shaped learning and frequency effects in a multilayered perceptron: implications for child language acquisition', *Cognition*, vol.38, no.1, pp.43–102.
- Plunkett, K. and Marchman, V. (1996) 'Learning from a connectionist model of the acquisition of the English past tense', *Cognition*, vol.61, no.3, pp.299–308.
- Prince, A. and Smolensky, P. (1997) 'Optimality: from neural networks to universal grammar', *Science*, vol.275, no.14, pp.1604–10.
- Rabinowitz, M. and Goldberg, N. (1995) 'Evaluating the structure-process hypothesis' in Weinert, F.E. and Schneider, W. (eds) *Memory Performance and Competencies: Issues in Growth and Development*, Hillsdale, NJ, Lawrence Erlbaum.
- Richards, D.D. and Siegler, R.S. (1981) 'U-shaped curves: it's not whether you're right or wrong, it's why' in Strauss, S. and Stavy, R. (eds) *U-shaped Behavioural Growth*, New York, Academic Press.
- Roberts, S. and Pashler, H. (2000) 'How persuasive is a good fit? A comment on theory testing', *Psychological Review*, vol.107, pp.358–67.
- Rumelhart D.E. and McClelland, J.L. (1986) (eds) *Parallel-Distributed Processing: Explorations in the Microstructure of Cognition*, vol.1, Cambridge, MA, MIT Press.

- Salvucci, D.D. (2001) ‘Predicting the effects of in-car interface use on driver performance: an integrated model approach’, *International Journal of Human-Computer Studies*, vol.55, pp.85–107.
- Siegler, R.S. (1988) ‘Strategy choice procedures and the development of multiplication skill’, *Journal of Experimental Psychology: General*, vol.117, pp.258–75.
- Taatgen, N.A. and Anderson, J.R. (2002) ‘Why do children learn to say “broke”? A model of learning the past tense without feedback’, *Cognition*, vol.86, no.2, pp.123–55.

# Theoretical issues in cognitive psychology

## Chapter 17

Tony Stone

---

### 1 Introduction

In the preceding chapters you have met a wide variety of competing theories of our cognitive capacities. In addition to the debates between competing specific psychological theories (e.g. of visual perception, autobiographical memory or reasoning), cognitive psychology has seen debates that cut across the different areas of cognition. These debates concern the key concepts and explanatory strategies that are used in modelling cognition. This chapter introduces you to a selection of these debates.

#### ACTIVITY 17.1

---

Look through the notes relating to theories or models of cognition that you have made on previous chapters and try to identify concepts and themes that recur.

#### COMMENT

You might start by simply making a list of all the theories or models that have been discussed in the previous chapters and then look for similarities and differences between them. If you look, for example, at the DRC model of word recognition (Chapter 6), you might note that there are two routes for the pronunciation of written words, one route that involves assembling the phonological representation of the word using letter-sound rules (also called the 'rule-based' route in Chapter 6), and another – addressed route – linking the written form and the correct pronunciation (also called the 'lexical' route in Chapter 6). You might then investigate the extent to which other cognitive models involve these two distinct kinds of *mental processing*.

---

In doing Activity 17.1 you may have noticed that all of the cognitive theories discussed in this book (except, perhaps, Gibson's model of perception) make use of the notion of *mental representation* (in addition to the notion of *mental processing* mentioned in Activity 17.1). The Bruce and Young (1986) model of face recognition discussed in Chapter 4, for example, contains a component labelled *face recognition units* (FRUs): these are mental representations of the faces with which you are familiar. Similarly, models of spoken word recognition (Chapter 6) make use of the notion of a *mental lexicon* – this can be thought of as a store that contains mental representations of all the words that you know and representations of various properties of those words (e.g. their meaning). In the study of memory, there are thought to be mental representations of the episodes in one's life (in episodic memory) and of the 'know how' that one has (in procedural memory). In each of these examples, the idea is that there is something internal to the mind that encodes information about the world, one's knowledge of the world, or of a past, present or future experience.

A fundamental principle of contemporary cognitive psychology is that cognition involves *the processing or transformation of structured mental representations*. There are two basic kinds of processing in evidence in contemporary cognitive models. Sometimes representations are transformed by *rules*. For example, in Marr's model of the early stages of visual processing (see Chapter 3) we find the successive transformation of mental representations by a kind of rule called an algorithm. Thus, the grey level description is transformed into the raw primal sketch by a smoothing algorithm. Similarly, in the DRC model of word reading, the pronunciation of non-words such as SLINT 'requires the postulation of a system of rules ... loosely referred to as the letter-sound rule system' (Coltheart *et al.*, 2000, p.370; first published 1993). On other occasions mental processing is conceptualized as the *transmission of activation* from one representation or group of representations to others. Whilst this kind of mental processing is particularly characteristic of connectionist models of cognition (see Chapter 16), it is also found in more traditional models. In the Bruce and Young model of face recognition, there are connections between the FRUs and the person identity nodes (PINs) with activation flowing from an active FRU to the associated PIN.

The idea that cognition can be understood in terms of the rule-guided transformation of mental representations is at the heart of the computational model of the mind that has been dominant throughout most of the history of cognitive psychology (though this is often coupled with transmission of activation). The first theoretical debate that we shall examine, in Section 2, concerns whether human information processing involves rules and representations to the extent proposed by many of the cognitive models you have studied.

A further thing you may have noticed when doing Activity 17.1, though you might have thought it too obvious to mention, is that there are different models for different cognitive competencies. Thus, there are separate models for spoken word recognition and visual word recognition (Chapter 6), for object recognition and face recognition (Chapters 3 and 4), for episodic memory and for autobiographical memory (Chapters 8 and 14), and for speech production and comprehension (Chapters 6 and 7). This is not just a matter of practical convenience reflecting specialization of interest. It reflects the idea that the mind itself is composed of independent, special-purpose systems that do specific information-processing tasks – i.e. that the mind is modular. Section 3 of the chapter is devoted to this debate.

These two debates have been central to recent theoretical work, but there are other important questions that have also been subject to intense discussion. For instance, it is one of the attractions of many connectionist models that they are (allegedly) more brain-like than traditional models. Yet it is noticeable that very little is said in the previous chapters about the way in which the various models of cognitive functioning are actually implemented in the brain – there is little connection, seemingly, between the activity of cognitive modelling and the actual study of the physical brain. Nonetheless, much of the most compelling data for cognitive models over the past 30 years or so has come from the study of people who suffer from various kinds of cognitive impairment as a result of brain damage (Ellis and Young, 1998, give a textbook introduction to this work). For instance, and to take just one example from many, the study of people with prosopagnosia

(a specific inability to recognize once familiar faces) has provided important evidence for the development of the Bruce and Young model of face recognition. This interaction between the study of normal and disordered cognition raises theoretical questions. For example, how exactly are inferences about normal function made from the study of single cases of disordered cognitive functioning? These questions are taken up in Section 2, in the context of the debate about rule-guided mental processing.

The importance of neuropsychological evidence also highlights the need for a clear account to be given of the relationship between the kinds of models discussed in this book and the study of the brain. Will it be the case, for instance, that ultimately cognitive psychology will reduce to (and perhaps be replaced by) cognitive neuroscience? These issues will be the topic of Section 4.

## Summary of Section 1

---

- A fundamental principle of contemporary cognitive psychology is that cognition involves the processing of structured mental representations.
- Two kinds of processing are in evidence in cognitive models: rule-guided and the transmission of activation.
- Cognitive modelling reflects the idea that mind is composed of modular systems.
- Inferences are made from data from neuropsychological case studies to models of normal cognitive functioning.
- There is a need to clarify the relationship between cognitive models and neurobiological models of the brain.

## 2 Computation and cognition

### 2.1 Some basic ideas

One question that might occur to you as you come to the end of this book is whether there is a *general theory* of the mind (or of mental functioning) that can be distilled from the preceding chapters. After all, one of the aims of science is to try to find general theories for seemingly disparate phenomena. An idea that has been influential throughout the history of cognitive psychology – indeed that was present at its inception – is the idea that the mind is a computational device: that cognition is computation. We have already met the basic ideas that lie behind this approach in discussing Activity 17.1 – it is the approach that sees cognition as a matter of the rule-guided transformation of structured mental representations. I refer to this idea as the **computational model of the mind (CMM)**.

An excellent example of these ideas in action is the DRC model of word reading that you met in Chapter 6, and that was mentioned in Section 1. This model proposes that there is an assembled phonology route where the pronunciation of regular words

and non-words is computed via letter-sound rules. The pronunciation of the non-word SLINT, for instance, is arrived at through the application of the rules that S is pronounced /s/ L is pronounced /l/, I is pronounced /ɪ/, and so on. It is important to be clear that these letter-sound rules are meant to be part of the *causal* story of how regular words and non-words of English are pronounced. (Strictly speaking, familiar regular words will also generally be pronounced via the addressed route, but they can also be pronounced via these letter-sound rules.) The letter-sound rules are not thought to be simply *descriptions* of the way English regular and non-words happen to be pronounced. It is useful to compare the role played by letter-sound rules in the DRC with the role played by a rule that merely describes the operation of a system. Consider, for example, Ohm's law (this example is taken from Gallistel, 2001). Ohm's law states that in any electrical circuit  $I = \frac{V}{R}$  (where 'I' represents current, 'V' represents voltage, and 'R' represents resistance). The symbols 'I', 'V' and 'R' refer to measurable properties of an electrical circuit, and we can manipulate the symbols and make predictions that we can then test by measurement. For example, we can deduce from  $I = \frac{V}{R}$  that  $IR = V$ , and then we can measure current and resistance, compute the numerical product ( $I \times R$ ) and see if the number obtained is the same as that we find when we measure voltage (V). Ohm's law accurately *describes* the behaviour of electrical circuits, but it is not part of the *causal* story of why an electrical circuit behaves in the way it does. The electrical circuit itself does not contain *representations* of the current, voltage and resistance, nor a representation of the law describing their relationship.

It is essential to the CMM that cognition is computation in the sense that, in central cases, it involves the *rule-guided processing* of structured mental representations.

Why is this approach called the *computational* model of the mind? The reason is that the digital computer is an example of a physical device that can process information by transforming symbols via a program – a set of rules stored in memory. Thus, the thought lying at the root of the computational view is that the brain (a physical device) processes information in the way that a computer does in so far as human information processing involves transformation of mental representations that is guided by rules (akin to the computer program). It is important to recognize that this is a *model* of the mind. Models aim to capture what is fundamental to the thing being modelled – just as a geographical map (another kind of representation) attempts to capture what is fundamental to the terrain being mapped. So the CMM only aims to capture those aspects of the mind that are thought to be fundamental to information processing. Thus, the fact that a computer is made of silicon, wires, metal and plastic (the computer hardware), whereas human information processors are made of flesh and blood (our hardware or, perhaps, wetware), is not important, because these aspects of a computer are not part of the model (and reasonably so given the aim is to model the human mind, not human tissue).

But, you might wonder, shouldn't a model of the mind at least try to model that part of the human body – the brain – wherein mental processing occurs? Shouldn't we model the physical networks of neurons and the transmission of physical electro-chemical signals that actually implement information processing? It is a key aspect of the CMM that it does not try to model those processes – it does aim to provide a

model of the brain, but at a level that is more abstract than the physical level. As Ned Block (1995) puts it: the CMM models ‘the mind as the software of the brain’. It is a matter of debate, of course, whether or not the failure to model the brain’s physical processes is a virtue of the CMM or not.<sup>1</sup>

## 2.2 Connectionism versus the CMM: the past-tense debate

### 2.2.1 Connectionist modelling

I turn now to the first of the theoretical debates that I want to consider – connectionist objections to the CMM. Connectionism (also known as ‘PDP’ and ‘neural network modelling’) is a style of psychological theorizing and model building that re-emerged in the middle of the 1980s, and has become massively influential in cognitive psychology since then. You were provided with brief introductions to connectionism in Chapters 1 and 16, and examples of connectionist models have cropped up from time to time throughout the book.

A characteristic connectionist model is composed of three layers of artificial neurons or nodes – an input layer and an output layer with a layer of hidden units sandwiched between them. There are multiple connections between each of the layers that transmit activation from the input layer to the output layer via the hidden units. The level of activation passed from one node to another is a function of the activation of the nodes passing on activation and the weight of the connections between these nodes and the nodes that receive activation. Such models are trained to associate patterns of activation across the input layer with patterns of activation across the output layer. The patterns of activation in the input layer might, for instance, be taken to be representations of written words of English, and the output pattern of activation that the model is trained to produce taken to be representations of the pronunciation of those words. This would be a network whose target domain is the reading aloud of English words. Decisions have to be made by the modeller about how the patterns of activation across the input and the output layers represent the target domain. Roughly, the patterns found in the target domain have to be mirrored in some way in the patterns of activation in the input layer and by those, produced after training is complete, in the output layer.

A key aspect of these kinds of model is that they can learn to associate patterns; they are not programmed with an algorithm or rule that specifies what pattern should be associated with what. (There are various learning procedures, but we need not go into the details here.)

At their most radical (e.g. Seidenberg and MacDonald, 1999; McClelland and Patterson, 2002a and b), connectionist modellers aim to model human cognition in ways that dispense with rule-governed mental processing (sometimes connectionist modellers say that they aim to provide a different conception of rule-guided

---

<sup>1</sup> It is often thought that the CMM must be wrong because we are conscious and computers are not. But this is too quick. The CMM is a model and does not have to capture all aspects of human psychology. Perhaps a different model could deal with consciousness. However, it would be an objection to the CMM if failing to model consciousness meant that it thereby failed to model cognition, an objection made by John Searle (e.g. 1992).

processing, but this complication is left to one side). For instance, McClelland and Patterson say that, in the connectionist approach to the psychology of language, ‘cognitive processes are seen as graded, probabilistic, interactive, context-sensitive and domain-general. ... Characterizations of performance as “rule-governed” are viewed as approximate descriptions of patterns of language use: no actual rules operate in the processing of language’ (McCLELLAND and Patterson, 2002b, p.465).

Connectionism thus attacks a fundamental aspect of the CMM. It claims that mental processing is typically not rule-guided.

### 2.2.2 The past-tense debate: the words and rules model

I am going to present the connectionist challenge to rule-guided mental processing by considering the so-called past-tense debate (Pinker and Ullman, 2002a and b; McClelland and Patterson, 2002a and b). In this subsection, I set the scene for the challenge by first describing an approach to the debate that does conform to the CMM.

The past-tense debate concerns how we should best understand the ability, possessed by all competent speakers of English, to form the past tense of English verbs (the debate has also been concerned with how children develop the ability to form the past tense, but I leave such developmental questions to one side in this chapter). This example is not as complex as some of the models and tasks you have met in this book, but it has been the subject of considerable debate, and the simplicity of the phenomena being modelled allows one of the fundamental divides between CMM and connectionism to be clearly seen.

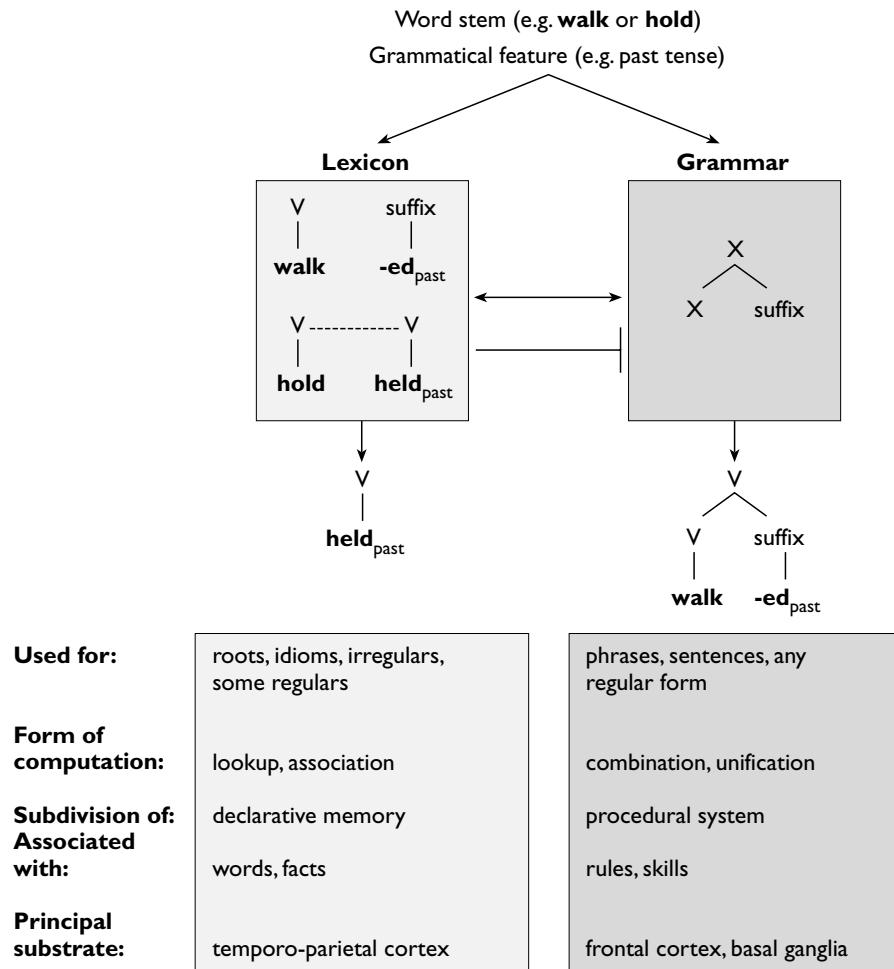
Consider, then, the problem of how a competent speaker of English puts a verb into the past tense. The past tense of the overwhelming majority of English verbs takes the form VERB STEM + PAST TENSE MORPHEME<sup>2</sup>. Thus the past tense of the verb TO HUNT is HUNT + ED (HUNTED), the past tense of the verb TO STROLL is STROLL + ED (STROLLED), the past tense of the verb TO JUMP is JUMP + ED (JUMPED), and so on. However, there is a minority of verbs (around 160 of the most common in the English language) that do not follow this pattern. These verbs form their past tense in a seemingly irregular fashion. Thus the past tense of the verb TO GO is WENT (not GO + ED) and the past tense of the verb TO BUY is BOUGHT (not BUY + ED).

One way to model how this task is accomplished has been developed by Steven Pinker and his colleagues (e.g. Pinker and Ullman, 2002a and b; Ullman *et al.*, 1997). This model – the **words and rules model** – ‘claims that the regular-irregular distinction is an epiphenomenon of the design of the human language faculty’ (Pinker and Ullman, 2002a, p.456). In outline, this model posits that two separate structures of the language faculty are responsible for the formation of the past tense: the lexicon and the grammar. When a past tense is to be formed, both the grammar and the lexicon are accessed in parallel. Verbs that form an irregular

---

<sup>2</sup> This is a simplification, since the phonology of the past tense morpheme does vary (compare the pronunciation of the past tense morpheme in HUNTED, WISHED, and STROLLED, for example).

past tense will access the appropriate form in the lexicon. Verbs that form the regular past tense will access no past tense form in the lexicon, and the grammar component of the language faculty will add the regular ending or inflection. (This is, then, a dual-route model analogous to the DRC model with which you are familiar.) Given that the lexicon and the grammar are accessed in parallel, the model posits an inhibitory signal from the lexicon to the grammar. Whenever an irregular past tense is formed this signal blocks the parallel formation of an incorrect regular form. This functional model has been coupled with the



**Figure 17.1** A simplified illustration of the words and rules theory and the declarative/procedural hypothesis. When a word must be inflected, the lexicon and grammar are accessed in parallel. If an inflected form for a verb (V) exists in memory, as with irregulars (e.g. held), it will be retrieved; a signal indicating a match blocks the operation of the grammatical suffixation process via an inhibitory link from lexicon to grammar, preventing the generation of *holted*. If no inflected form is matched, the grammatical processor concatenates the appropriate suffix with the stem, generating a regular form

Source: Pinker and Ullman, 2002a, Figure 1, p.457

declarative/procedural hypothesis that proposes that the lexicon is part of declarative memory and is subserved by temporal and temporo-parietal regions of the neocortex, and that the grammar component is part of procedural memory and is subserved by the basal ganglia and the areas of frontal cortex to which they project. This proposal melds the words and rules theory onto structures known from independent evidence to be involved in language processing. Moreover, independently of the debate about the past tense, declarative memory is thought to be responsible for the retention of facts, and procedural memory responsible for the learning and control of motor skills, including skills that require sequencing. The thought, then, is that retention of an irregular past tense is the retention of a fact, whereas the construction of regular past tense forms is retention of a procedure.

### 2.2.3 Connectionist modelling of the past tense

#### ACTIVITY 17.2

---

Try to identify the aspects of connectionist modelling that make it different from many of the other kinds of psychological models you have met in this book. For each difference try to find a specific connectionist model that illustrates the difference. In addition to this book, a good – freely available – web-based resource that will help you is the entry on connectionism (Garson, 2002) in the *Stanford Encyclopedia of Philosophy* (<http://www.seop.leeds.ac.uk/entries/connectionism/>)

#### COMMENT

One difference you might find mentioned is that connectionist models are more 'brain-like' than models that are characteristic of the CMM. In what ways are connectionist models more 'brain-like' and are they very 'brain-like'?

---

Connectionist modelling aims to model cognitive abilities without relying on rule-guided mental processing. Let's look at how it does this with respect to the past tense. The first connectionist model of past-tense formation was developed by Rumelhart and McClelland (1986). This was a simple two-layer network that was trained to associate the correct past-tense form to both regular and irregular English verbs. Yet it did this with just a single route. All verbs, both regular and irregular, were processed via the same set of units and connections. This model appears to show that connectionist models are capable, in principle, of accounting for what looks like rule-guided behaviour in a way that does not require the use of a rule!

Rumelhart and McClelland's model was a simple pattern associator. It learned to associate the appropriate past-tense forms with the appropriate verb stem by relying on the statistical regularities contained in the training data. No rule was programmed into the model, and the trained model did not learn by forming its own rule, such as 'add ED to regulars'. How then does the model do it? Understanding this – and the same goes for any connectionist model – requires exploring in detail the node activations and connection strengths between nodes in the trained network. It should not be assumed, for instance, that the Rumelhart and McClelland model forms the past tense of all verbs in the way that the words and rules model forms the past tense of irregular verbs. It should not be assumed, that is, that the correct phonology of the

past-tense form is stored and then addressed by the appropriate verb stem – that there is simply a list of verb stems and their past-tense forms that are linked one-to-one with each other.

The Rumelhart and McClelland model came in for some severe criticism from defenders of traditional rule-based theories (e.g. Pinker and Prince, 1988). However, later models were able to meet many of these criticisms (e.g. Plunkett and Marchman, 1993). What Rumelhart and McClelland (and other connectionist modellers) are thought to have shown is that there is an alternative to dual-route, words and rules style models. But do these connectionist models provide a better account of the available data than more traditional types of model?

McClelland and Patterson (2002b) argue that connectionist models do provide a better fit with the data. They argue that the English past tense is a *quasi-regular* domain. They point out that past-tense forms of English seem to fall into eight groups. For instance, there is a group of verbs, including SAY, DO, TELL, SELL and FLEE, whose past tenses are SAID, DID, TOLD, SOLD and FLED. These verbs form a cluster that forms the past tense by adding /d/ with a vowel adjustment to the stem. Another group, including BRING, CATCH, SEEK, TEACH and THINK, have the past tenses BROUGHT, CAUGHT, etc. They form their past tenses by replacing the final consonant cluster with /t/ and adjusting the middle vowel to /ɔ/ (sounds like ‘aw’). McClelland and Patterson (2002a) claim that the words and rules model cannot capture these quasi-regularities but that they can be captured by connectionist models that have a single system for the formation of the past tense. They explain that their network can make the transformation KEEP → KEPT by simply

adjust[ing] the activations of the output units representing the vowel, something the network will have learned to do on the basis of experience with *keep* and its neighbours *creep*, *leap*, *sleep*, *sweep*, and *weep*. The network uses the same connection-based knowledge that allows it to perform the regular mapping, and also taps into specific connections activated by the particular properties of *keep* to produce the vowel adjustment.

(McClelland and Patterson, 2002a, p.464)

Supporters of the words and rules model of the past tense are not without responses. Pinker and Ullman (2002a) draw attention to other aspects of past-tense formation that they believe will be hard for connectionist models to deal with since their linguistic explanation lies in relatively deep linguistic principles. An example they provide is the way in which some usually irregular past-tense forms (e.g. RING → RANG and STAND → STOOD) become regular in certain contexts such as RINGED THE CITY and GRANDSTANDED. These forms (regularizations of irregulars) occur due to linguistic principles concerning the formation of complex words of English. TO RING and TO GRANDSTAND are verbs that are derived from nouns (A RING and A GRANDSTAND). Because of this, it is not possible for the irregular form (RANG/STOOD) stored in memory to be accessed – since that form must be accessed via a verb stem. Hence, the regular rule kicks in since no inhibitory block is sent from the lexicon to the grammar.

As of now, the empirical evidence appears to provide no clear way to adjudicate between these two accounts. It would be a brave theorist, however, who predicted that connectionist models will prove *unable* to handle the data presented by words and rules theorists, given the good track record that connectionist modelling has of responding to these kinds of challenges. However, it should also be pointed out that – as grammatical phenomena go – the past tense of English is relatively straightforward. The fate of the rival views of mental computation is not going to be sealed by how this one debate turns out. The grammars of language are extremely complex and connectionist modellers have much work to do to show that CMM-inspired models are inadequate. So, the jury is still out – indeed, it would be more accurate to say that it will be some considerable time before the jury can even retire to consider its verdict!

#### **2.2.4 Using evidence from cognitive neuropsychology to inform the past-tense debate**

In this subsection I consider how evidence from the study of people with acquired linguistic impairments can be used to inform the past-tense debate. Whilst this is of intrinsic interest it will also allow us to discuss another important theoretical question: how are inferences about normal cognitive functioning made from cognitive neuropsychological case studies? Evidence from neuropsychological case studies has been a major influence on the development of many of the psychological models you have met in previous chapters. We would do well, therefore, to understand how data from case studies are used in the development of models of normal function.

##### *Neuropsychological evidence on the formation of the past tense of English*

Marslen-Wilson and Tyler (1998) investigated the pattern of linguistic abilities and impairments of three people with agrammatism (i.e. difficulty in comprehending and producing inflected forms of English). Their investigation used what is called the primed lexical decision task. In this task a target word is preceded by a prime that is either a morphologically related word or a semantically related word (e.g. ‘jumped-jump’ or ‘swan–goose’). In people whose language abilities are unimpaired, there is a faster response to a target word that has been primed than to one that has not in *both* these conditions. The key question that Marslen-Wilson and Tyler asked is whether priming occurs for their language-impaired participants and, crucially, whether it occurs for both regular and irregular past-tense primes (e.g. whether JUMP primes JUMPED and BUY primes BOUGHT). Two of the participants showed a positive priming effect for the irregular past-tense forms, but not for the regular past-tense forms (i.e. BUY primed BOUGHT, but JUMP did not prime JUMPED). A third, however, showed the opposite pattern of response and was positively primed by regular past-tense forms but not by irregular forms. Marslen-Wilson and Tyler argue that this is good evidence that two different kinds of mental computation are required in order to form the past tense of English – a rule-guided computation (for the regular past-tense forms) and an associative link computation (for the irregular past-tense forms). Why else, one might reason, would one find this particular pattern of performance? Marslen-Wilson and Tyler’s

findings are an excellent example of what is called a **double dissociation** of psychological functioning, which is often taken to be very good evidence for the existence of separate processing routes – in this case, separate routes for the processing of regular and irregular past-tense forms. Why is it *especially* good evidence?

*Drawing inferences about normal functioning from psychological impairments*

We can answer this question step-wise by considering the following questions:

- 1 Why is evidence from brain-damaged people relevant to models of normal functioning at all?
- 2 Why is evidence from a double dissociation thought to be especially good evidence for the existence of separate processing routes? Specifically, why does it provide better evidence than a single dissociation?
- 3 If impairments can be dissociated, presumably they can be associated. For instance, what inferences could we draw from performance on the task described above where the patient was primed neither by regular nor irregular primes?

*Response to Question 1* There is no a priori reason to believe that evidence from brain-damaged people will or will not be relevant to normal functioning. This is an empirical issue. However, the *modular* nature of models of normal cognitive functioning suggests that this might be so. It only *suggests* this, however, since the brain may turn out to be organized in such a way that damage to one part of the brain often, or always, results in seemingly random patterns of impairment that fail to line up in any way at all with models of normal functioning. But it is a consistent neuropsychological finding, from the first such investigations at the end of the nineteenth century until the present, that brain injury often results in damage to specific cognitive functions as modular cognitive models suggest and not in random patterns of impairments. Another consistent finding is that the functions left unimpaired either work as normal (often called ‘the locality assumption’) or, if not normally, at least in ways that can be understood in terms of normal processing (often called the ‘transparency assumption’) (Johnston and Braisby, 2000).

*Response to Question 2* Schematically, a double dissociation occurs when patient A is impaired in their performance on psychological task 1, but performs at normal or near normal level on psychological task 2 (single dissociation), *and* patient B is impaired on psychological task 2 but unimpaired or near normal level on psychological task 1 (second single dissociation). The double dissociation provides strong evidence that the two tasks involve separate information-processing routes or mechanisms. A single dissociation does not allow us to draw this conclusion – it could be argued that poor performance on the impaired task is simply due to that task being the more difficult of the two.

*Response to Question 3* Consider a patient who has impaired performance on task 1 and on task 2 – at the limit, a patient whose ability on both tasks has been completely destroyed. Would this be good evidence that a single route or processing mechanism was responsible for both tasks? It would not. The reason it would not is

that an *association* of impairments may arise for reasons that have nothing to do with how the human cognitive system is organized. Specifically, it might arise due to neuroanatomical accident. Brain damage is often devastating, leading to multiple impairments and not to the *specific* impairments we are considering. But this is often due simply to the extent of the physical damage. Moreover, even in the case of an association of impairments between two tasks in the same psychological domain (such as past-tense formation), the inference to a common processing mechanism would be unwarranted. The association of cognitive impairments may be due to the fact that the brain areas that subserve the two different processing mechanisms are both damaged, either because of the extent of damage, or because, even though the damage is localized to a relatively small region of the brain, that small region just happens to subserve both mechanisms. Perhaps the mechanisms are adjacent to one another in the brain. Or perhaps the blood vessels in the brain are so arranged that the brain regions that subserve the two mechanisms are both supplied by the same blood vessel. Damage to that blood vessel might then result in associated impairments. In this latter case, it's particularly easy to see why the inference from an association of impairments to a common information-processing mechanism should not go through – surely our psychological models don't have to answer to the arrangement of blood vessels in the brain?

#### *The connectionist response*

If all of the above is correct, do data such as those presented by Marslen-Wilson and Tyler refute the connectionist position on the past tense or, if not refute it, at least provide compelling evidence in favour of the alternative words and rules type approach? Supporters of this type of approach think that it does, of course. But connectionist modellers have attempted to show that, at the very least, these data are not *conclusive*. They have done this in two related ways. First, they have argued that, in principle, double dissociations of function might be found even if there is a single processing mechanism that is responsible for performance on two tasks. Second, they have built single-route connectionist models that – when ‘lesioned’ – can simulate the patient data (Juola and Plunkett, 2000, present a connectionist model that simulates the Marslen-Wilson and Tyler data).

The claim that double dissociations don't logically guarantee the existence of two routes or processing mechanisms responsible for the tasks in question is certainly correct (Chater and Ganis, 1991). But scientific theories or models are not the logical deductive consequences of data. The relationship between theory and evidence is *non-demonstrative*. Indeed, it is abductive – i.e. a matter of inference to the best explanation. Suppose a cognitive neuropsychologist has some data D, perhaps a pattern of impairment across a number of patients that constitutes a double dissociation. She then proposes a model – M1 – involving two routes or processing mechanisms that accounts for the data. The claim can only be that this model provides a good, or, with fortune, the best explanation for that data. If another model – M2 – is proposed that accounts for the data and implicates only a single route or processing mechanism, then that is a *competing* explanation. In order to decide between the two models we have to ask questions such as: are there other data that M1 can account for that M2 can't, or vice versa? Or, is one of the models implausible on independent grounds? And, of course, the same logic applies if the competing model, M2, is a connectionist model involving a single route that can simulate the

patient data. In sum, double dissociations *are* compelling pieces of evidence for models that propose separate processing routes or mechanisms – but they are not conclusive (Coltheart and Davies, 2003 give a clear and trenchant presentation of this case).

### 2.2.5 Rules in connectionist networks

If connectionist modelling is to be a genuine alternative to the CMM, then it is vital that the models do not embody knowledge of rules or, at least, that they do not embody knowledge of rules in the same sense that models conforming to the CMM do. In a series of papers, the philosopher Martin Davies (e.g. 1990a, 1990b, 1995)<sup>3</sup> has shown how certain kinds of connectionist architecture do embody knowledge of rules in the sense that the rules figure in the causal story of how the model works and not just as a description of the model's performance. (Remember the distinction drawn in Section 2.1 between the *causal* role played by the letter-sound rules in the DRC model of reading and the role of Ohm's law in the accurate *description* of the behaviour of an electronic circuit.)

Davies distinguishes a number of different things that might be meant when it is said that a cognitive system or model embodies knowledge of a rule. I will explain just two of the notions that Davies discusses. The first sense is when a rule is *explicitly* encoded or represented in a system or model. An example of such a case is, again, the implementation of the assembled phonology route in the DRC model of reading. Coltheart *et al.* (2000) describe the operation of the letter-sound rules as follows:

When confronted with a letter string for translation, it seeks to apply the rules to the string from left to right, starting with the longest possible rule that could accommodate that string. For the word *chip* it would start with four-letter rules, looking for a rule that maps the letters *chip* on to a single phoneme. No such rule will be found in the rule base. So a rule corresponding to the first three letters *chi* is sought; none will be found. The search for a rule for the first two letters *ch* [ $ch \rightarrow /tʃ/$  (sounds like 'ch')] will, however, be successful.

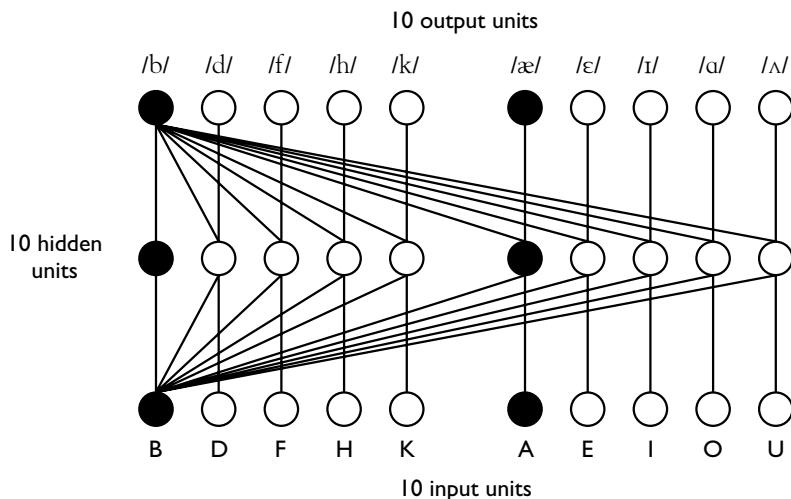
(Coltheart et al. 2000, p.392)

A second notion of knowledge of a rule that Davies considers comes from reflection on a 'toy' connectionist network that takes representations of a small number of written consonant–vowel pairs and associates these with a representation of their pronunciation. For example, when given a representation of the written pair BA as input it produces a representation of its pronunciation /bæ/ (sounds like 'ba' as in 'bat') as output. The network is represented in Figure 17.2 overleaf.

Davies argues that this network does indeed embody tacit knowledge of letter-sound rules, but clearly not in the sense that the rules are explicitly encoded. The

---

<sup>3</sup> This subsection is heavily indebted to Davies (1995), but I have drastically simplified his discussion.



**Figure 17.2** A connectionist model that embodies tacit knowledge of letter-sound rules

Source: based on Davies, 1995, Figure 3, p.181

sense in which it does embody tacit knowledge of the letter-sound rules relies on the concept of a *common causal explanation*. Davies' idea is that if there is in the model or system

a component mechanism, or processor, or module that operates as a *causal common factor* to mediate all the input-output transitions that instantiate the pattern described by the rule ... then the rule is said to be implicit in the system (or the system is said to have implicit or tacit knowledge of the rule).

(Davies, 1995, p.162)

Examine the toy connectionist model shown in Figure 17.2 carefully. Does it have a common causal factor in this sense? The point to notice is that whenever a consonant-vowel letter string containing B is input into the system – regardless of what vowel it is concatenated with – the same unit represents that letter (i.e. B) and the same phonological output unit (i.e. /b/) is activated. Thus we can say of this model that the connection between the unit for B and the unit for /b/ will always be active in the pronunciation of a B\_string, and hence is a causal common factor in the processing of all B\_strings. The same, of course, goes for all the other connections. The letter-sound rules that capture the pronunciations of these two-letter strings are, therefore, implicit in the system.

The point to take away from this discussion is not that all connectionist models embody implicit knowledge of rules in this second sense, just that some may. When examining a connectionist model that claims to be dispensing with implicit rules, you need to ask yourself whether this is really what it has done. And if it doesn't embody rules in either of the ways just discussed, then you need to try to discover how the model does complete the task set for it.

## 2.2.6 Connectionism, structure and compositionality

Connectionist modelling also challenges the style of *mental representation* that is characteristic of the CMM. This is a complex and (in parts) technical debate so I am only going to discuss one issue briefly. (Interested readers should consult Crane, 2003 and Clark, 2001 for introductory treatments of this challenge. The seminal work is that of Fodor, e.g. 1975 and 1987, Fodor and Pylyshyn, 1988, and the papers collected in MacDonald and MacDonald, 1995.)

CMM holds that the mental representations processed by rules must have *compositional structure*. This explains a fundamental property of our cognitive systems – its **systematicity**. This property can most easily be explained by reference to our linguistic abilities. Suppose I understand the sentences ‘The dog bit the cat’ and ‘The cow jumped over the moon’. My understanding is systematic because I need to learn nothing more in order to understand the sentences ‘The cow bit the dog’ and ‘The cat jumped over the moon’. Sentences are composed of words, and the same words can be deployed in different sentences to express new thoughts. It is this **compositionality** of language that explains its systematicity. Contrast this with what might happen when someone learns a second language from a phrase book. It is perfectly possible for a monolingual English speaker to learn to say the Romanian sentences ‘O sticle cu vin roşu’ and ‘Un caiine cu dinţi mici’ in appropriate circumstances and yet be unable to combine the words in these sentences in new ways to express appropriate new thoughts. You could only do this if you knew what each word means and have some rudimentary grasp of the grammar of the language – and knowledge of the meaning of whole phrases or sentences doesn’t guarantee that you will gain this (indeed typically you will not). The CMM supposes that much of our cognitive processing involves the processing of mental representations that have compositional structure.

Connectionist models challenge this approach to mental representation in that it is often claimed that these models can explain the systematicity of human cognition without using compositional mental representations (at least, in the sense of compositionality used by the CMM).

It is certainly true that connectionist models can be so designed that they capture the systematicity of human cognition. The issue turns, however, on how they do this. One way in which they might do so is by simply implementing a compositional representational system. But, if that is how connectionism captures systematicity, then it offers no challenge to the CMM’s account of mental representation. So, how might a connectionist model capture systematicity *without* using compositional mental representations? Consider this simple example – this time in the domain of thought, not language. It seems to be a fact about human thought that if I am able to think the thought *New York is dangerous* and to think the thought that *London is safe*, then I can also think the thought *New York is safe* and the thought *London is dangerous* (systematicity again). The CMM, of course, explains this by saying that thought (like language) is compositional. A connectionist model might capture this systematicity by having four separate bunches of nodes representing each of these four thoughts. These nodes could then be trained so that activation of the bunch of nodes representing *New York is dangerous* and of those representing *London is safe* results in the activation of the two sets of nodes representing the thoughts *New York is safe* and *London is dangerous*. This would, indeed, be systematicity without

compositionality. But the advocate of the CMM can reply that whilst the compositional style of representation used by the CMM guarantees the systematicity of thought, the representational system in the supposed connectionist model does not. The connectionist model, it might be said, only captures systematicity by *accident* of the learning regime involved. That is, the network has to be appropriately trained so as to ensure the appropriate transitions. Just being able to entertain the first two New York and London thoughts doesn't guarantee being able to entertain the second two New York and London thoughts. But a thinker of thoughts is so guaranteed, so non-compositional representational systems are inadequate. Connectionists are not without reply, of course.

## Summary of Section 2

---

- The CMM is the view that cognition is computation.
- The computational model of the mind (CMM) sees mental processing as involving the rule-guided transformation of structured mental representations.
- Connectionist models are computational models that challenge the idea that mental processing is rule-guided. In connectionist models the primary role is played by the transmission of activation.
- The dual-route words and rules model illustrates the CMM approach. Connectionist modellers have demonstrated that models involving a single route can be trained to form the past tense of verbs via the transmission of activation.
- Neuropsychological evidence of double dissociation of impairment from case studies of patients with grammatical impairments presents compelling but not decisive evidence in favour of the CMM approach.
- There is a question as to whether or not connectionist models are really rule-free. Whilst they do not explicitly encode or represent rules, careful investigation is needed to check whether they encode rules implicitly.
- Connectionist models also challenge the need for mental representations to be structured compositionally.

## 3 Modularity

### 3.1 An outline of Fodor's theory of modularity

Modularity should have been near the top of the list of concepts and themes relating to theories or models of cognition that you identified in Activity 17.1. But exactly what is meant when a psychological system or function is said to be modular? One view that has been taken by cognitive psychologists is that we can give an *operational definition* of modularity. The basic idea of an operational definition is to define a theoretical concept in terms of the operations used to measure them (thus, intelligence might be said to be operationally defined by the tests used to measure it.)

For instance, Tim Shallice quotes and endorses a definition of modularity proposed by Endel Tulving who suggested that two psychological systems are functionally different where:

One system can operate independently of the other although not necessarily as efficiently as it could with the support of the other intact system. The operations of one system could be enhanced without a similar effect on the operation of the other; similarly the operations of one system could be suppressed without a comparable effect on the activity of the other. The functional difference also implies that in important, or at least non-negligible ways, the systems operate differently, that is that their function is governed at least partially by different principles.

*(Tulving, quoted in Shallice, 1988, p.21)*

Whilst Tulving's definition might be helpful for identifying modules, it's not clear that it gives us a particularly helpful characterization of what a modular system is. (Think: I might have a good way of identifying UFOs, but not know much about how they work.) Jerry Fodor (1983) has, however, provided us with a detailed discussion of modularity in his book *The Modularity of Mind*. In outline the account he gives is straightforward.

For Fodor, the mind is divided into three different types of system: (1) sensory transducers; (2) modular input systems; and (3) non-modular central systems.

The sensory transducers pick up physical stimuli from the environment – photons hitting the retina, sound waves causing the tympanic membrane to vibrate etc. – and transform these stimuli non-computationally into a format or code that the brain can understand. Recall that in Chapter 6, Section 2.1.1 you were provided with a picture of the waveform of a blast of speech. That picture is a representation of some of the physical properties of the speech stream. The sensory transducers have to transform these physical properties into a format the language processing system can understand.

The modular input systems mediate between the sensory transducers and the central systems. They provide, we might say, the perceptual experiences that provide a database of evidence for the processes of belief formation and decision making that are jobs for the non-modular central systems. The non-modular central systems also contain all the encyclopaedic knowledge one has stored in memory. In forming beliefs and taking decisions, the central systems will typically take account both of the evidence of the senses and stored knowledge.

Fodor characterizes the modular input systems as those that possess the following cluster of properties:

- Informational encapsulation
- Domain specificity
- Shallow output
- Mandatory operation
- Limited central access to the mental representations that modules compute
- Fast speed of operation

- Fixed neural architecture
- Characteristic and specific patterns of breakdown
- Characteristic pace and sequencing of their ontogeny.

Fodor does not say that a psychological system must possess *all* of these properties to be modular; indeed he explicitly states that his concept of modularity is a cluster concept and that modularity admits of degrees. However, it is clear from some of Fodor's other writings that some of these properties are more important to his conceptualization of a modular input system than are others (e.g. Fodor, 1985a). In particular, Fodor puts great importance on the properties of *domain specificity* and *informational encapsulation*, and these two properties have been the focus of much of the ensuing debate on Fodor's account of modularity. I will, therefore, limit my discussion to these two properties.

### 3.1.1 Domain specificity

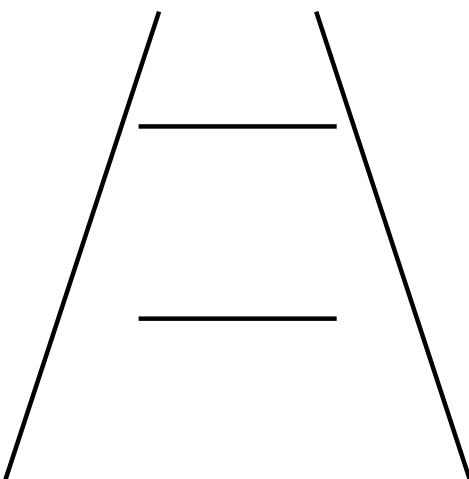
In *The Modularity of Mind*, Fodor's main examples of modular input systems are those responsible for visual perception and for the recognition of spoken language. The claim that these systems are **domain specific** amounts to the idea that these input systems deal only with a limited, idiosyncratic, range of stimuli. The language module, for example, only deals with linguistic input. Thus it processes spoken language, and not more general environmental sounds such as the sound of a bell tolling. But it also processes visual sign language. So you should note that the modular input systems do not correspond in a one-to-one way to the five senses.

In cognitive psychology, it is domain specificity that is most often associated with modular claims, although the modules proposed are often more fine-grained than those discussed by Fodor. Cognitive psychologists have proposed, for instance, that the visual system breaks down into three modular subsystems that have the special-purpose jobs of processing the domain of face stimuli, processing the domain of objects and processing written stimuli. Moreover, in cognitive psychological models modules can be nested within one another – for instance, the face-processing module might itself be broken down into sub-modules. This can be taken as wholly within the spirit of Fodor's account.

### 3.1.2 Informational encapsulation

Informational encapsulation is the property that is non-negotiable for Fodor as regards whether a system is granted the status of a module. We need, therefore, to be clear about what it involves.

The easiest way to do this is via an example. Look at Figure 17.3. Does the top horizontal line *look* longer than the bottom one? I would be surprised if it did not. But the lines are, in fact, of equal length (measure them if you want to check). Now look at the picture again. Do the two lines *now* look the same length? I would be very surprised, this time, if they did. Of course, though you don't believe what you now see, you still see exactly what you did before. This is an example of **informational encapsulation**. You (now) have some knowledge – about the lines in the Ponzo illusion being the same length – but that knowledge cannot affect the processing that takes place within the visual-input module. The visual-input module is, if you like, sealed off from that information. The general point is that the processing of a



**Figure 17.3** The Ponzo illusion

modular input system cannot be affected by information stored in the central systems, nor by information being processed by other modules. The Ponzo illusion illustrates the former. The latter can be illustrated by noting that what you see would not be affected by someone saying ‘The lines are the same length’ *when you are looking at the Ponzo illusion* (i.e. by concurrent processing in a language module).

In understanding informational encapsulation, it is important to see that Fodor does not deny that there can be top-down information flow *within* a module. The ban on top-down information flow is from central systems to modular input systems. This has been a major source of confusion about Fodor’s views on modularity, so it is worth spending a little more time discussing it.

In *The Modularity of Mind*, Fodor considers an objection to the idea that language comprehension is informationally encapsulated which stems from experiments on semantic context effects in lexical decision tasks. For instance, it is well established that speed of reaction in a lexical decision task can be increased if the target word is preceded by a semantically related prime. Thus the word BREAD can be primed by the word BUTTER. This might be taken as evidence that the language-input module is not informationally encapsulated, on the grounds that it is information from one’s general encyclopaedic knowledge that explains these effects. Roughly, it’s because you *know* that butter is often spread on bread that BUTTER primes BREAD. However, Fodor uses the results of Swinney (1979) that you have already met in Chapter 6 to question the inference from semantic priming to the rejection of informational encapsulation. Swinney found that recognition of the word SPY was primed by a context sentence about espionage that included the word BUGS. This seems to be in line with the use of context effects to question encapsulation, since my general knowledge tells me that bugs (i.e. secret radio transmitters) are often to be found in association with spies. But, you will recall, Swinney also found that the word ANT was primed by that very same sentence about espionage. But why should my general knowledge of insects be brought into play by a sentence about espionage?

According to Fodor, what accounts for these findings is not the top-down influence on language comprehension of knowledge stored in the central systems,

but the organization of the mental lexicon itself. Perhaps there are links between semantically associated words such that presentation of a word leads to the transmission of activation to its semantic associates. But since the lexicon is *internal* to the language-input system, context effects would not then provide evidence that information from the central systems is influencing the processing of the language-input module. As Fodor warns, ‘it makes a difference ... where the information comes from’ (Fodor, 1983, p.80). If the information does not come from outside the module it fails to be a counterexample to encapsulation.

You should note that this line of response requires being able to draw a clear and non-arbitrary line between modules and central systems. It is controversial whether Fodor has met this requirement.

### ACTIVITY 17.3

---

Fodor divides the mind into modular input systems and non-modular central systems. Which of the following do you think are modular input systems and which non-modular central systems? In each case give reasons for your choice.

- Object recognition
- Face recognition
- Spoken word recognition
- Autobiographical memory
- Encoding and retrieval systems
- Problem solving
- Reasoning
- Spoken language production
- Reading aloud
- Attention
- Working memory.

### COMMENT

This is not as straightforward as it might seem. The earlier stages of object recognition and face recognition certainly seem to be input systems. They are modular in that they are domain specific, but do you agree with Fodor that they are informationally encapsulated? Whilst problem solving seems to be a good candidate for a non-modular central system, many would argue that memory is modular in that there seem to be different systems for different kinds of memory (declarative/procedural/autobiographical) – but memory isn’t an input system. How should attention be classified? Is it an input system? Is it modular? And what is to be said about output systems? Fodor didn’t talk about output systems in *The Modularity of Mind*, but they are often taken to be modular. Do you agree? Consider, for example, one output function – reading a word aloud: which of the criteria for modularity does it meet?

---

### 3.2 The central systems

The modular input systems pass information to the non-modular central systems and, as we have seen, this is one-way information flow. The information computed by the modular input systems is encapsulated from knowledge stored in the central systems and also from the output of other input modules.

The main function of the central systems is to integrate information from input modules with information stored in the central systems in the service of belief formation and decision making. It would be a very unwise creature that formed beliefs on the basis of the output of just one input module, or ignored things that it already knew and believed. Central systems, Fodor argues, must be domain general (able to cope with the output of many different input modules) not domain specific. Fodor argues that the central systems are also unencapsulated – in settling on what to believe, or in making a decision, anything in principle that the organism knows might be relevant. Thus the operation of the central systems is relatively slow compared with the fast speed of operation of the input systems.

Fodor's model for this account of the central systems is the type of non-demonstrative inference that is characteristic of the confirmation of hypotheses in science. Scientific confirmation is, Fodor says, characterized by two properties that he dubs **isotropic** and **Quinean** (after the famous twentieth-century American philosopher Willard van Orman Quine, 1908–2000, who argued that our knowledge was organized into a holistic web of belief). He explains these notions as follows:

By saying that confirmation is isotropic, I mean that the facts relevant to confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical ... truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect.

*(Fodor, 1983, p.105)*

By saying that scientific confirmation is Quinean, I mean that the degree of confirmation assigned to a given hypothesis is sensitive to the entire belief system; as it were, the shape of our whole science bears on the epistemic status of each scientific hypothesis.

*(ibid., p.105)*

The difference between these two properties needs a little spelling out. Here is one way to do this. In the philosophy of science a distinction is often made between the context of the *discovery* of a scientific theory and the context of the *justification* of a theory. Isotropy relates to the former and Quineanism to the latter. Discovering or inventing new theories requires creativity such as the ability to draw analogies between radically different domains (e.g. the analogy between the solar system and the structure of the atom that was drawn in the early part of the twentieth century by Bohr). But deciding whether to accept a new theory requires that account also be taken of the body of established scientific doctrine.

Theory confirmation is, therefore, a conservative process. So scientific reasoning involves a delicate balance between creativity and conservatism – and maintaining this balance requires that scientific confirmation (taken to include both the context of discovery and the context of justification) be unencapsulated. So, to the degree that scientific confirmation is a good model for ordinary belief fixation, the central systems must be unencapsulated too. The point made at the beginning of this section suggests that there is a good fit between scientific confirmation and ordinary belief fixation: it would be an unwise creature that formed its beliefs on the basis of the information provided by only one input module and without taking account of stored encyclopaedic knowledge. The analogy to scientific confirmation suggests a wise creature would be isotropic and Quinean.

Is there empirical evidence that the central systems are unencapsulated? Is scientific confirmation a good model of ordinary belief formation? Unfortunately, according to Fodor, evidence on this is scanty ‘given the underdeveloped state of psychological theories of thought and problem solving’ (Fodor, 1983, p.112).

#### ACTIVITY 17.4

---

Reconsider Chapters 10 and 12 on problem solving and on reasoning. Is there any evidence provided there that supports or undermines Fodor’s views on the central systems? Is he right that psychological theories of thought and problem solving are underdeveloped?

---

As by his lights evidence is lacking, Fodor points to two difficulties in the development of theories of the central systems that ‘are just the sort we should expect to encounter if such processes are, in essential respects Quinean/isotropic rather than encapsulated’ (1983, p.112). The first concerns the difficulties posed for artificial intelligence by the, so-called, **frame problem**. This is, very roughly, the problem of how to build a robot that can update its beliefs about the world as a result of the action it takes. Consider, for example, a robot that is given the job of making a phone call to Mary:

Let’s assume the robot ‘knows’ it can get Mary’s number and proceeds to dial. So far, so good. But now, notice that commencing to dial has all sorts of direct and indirect effects on the state of the world (including, of course, the state of the robot), and some of these effects are ones the device needs to keep in mind for the guidance of its future actions and expectations. For example, when the dialling commences, the phone ceases to be free to outside calls; the robot’s fingers (or whatever) undergo appropriate alterations of spatial location; the dial tone cuts off and gets replaced by beeps; ... and so forth. Some (but, in principle, not all) such consequences are ones the robot must be designed to monitor since they are relevant to ‘updating’ beliefs upon which it may come to act. Well, *which* consequences?

(Fodor, 1983, p.113)

We, of course, do this everyday updating effortlessly. But AI researchers have found building a device that can do it intractable. Notice that the frame problem arises because belief updating seems to involve keeping track of, more or less, everything – thus the problem is caused by the unencapsulated nature of belief updating that is required when performing the task.

The second difficulty in developing theories of central systems is that there is, according to Fodor, no neuropsychology of thought, whereas there is a well-developed neuropsychology of the modular input systems. We find evidence of specific damage to the face-processing system, to the spoken word recognition system, to object recognition, to the assembled route in the word-reading system, and so on. But, says Fodor, we do not find specific damage to components of the central systems, and this suggests that the central systems do not have a fixed neural architecture – one of the cluster of properties that modular input systems are said to possess. But is Fodor right? What about the evidence that memory systems can be selectively impaired, or that the ability to do arithmetical calculation can be destroyed, or that attentional systems can be damaged, or that theory of mind can be impaired<sup>4</sup>? (Shallice, 1988; Ellis and Young, 1998; and Baron-Cohen *et al.*, 1999 provide evidence on all these impairments.) It seems that Fodor either has to deny these are central systems, or deny that the property of fixed neural architecture carries any weight in the argument, or show that all of these deficits are actually the result of damage to the input modules that feed into them. Do any of these possible Fodorian ripostes strike you as promising?

Fodor believes that, regardless of how things might appear, the prospects for a computational psychology of the central systems are bleak. Indeed, he proposes what ‘some day will come to be known as “Fodor’s First Law of the Non-Existence of Cognitive Science”’; namely, that ‘the more global (e.g. more isotropic) a cognitive process is, the less anybody understands it. Very global processes, like analogical reasoning, aren’t understood at all’ (Fodor, 1983, p.107).

According to Fodor, the CMM just will not do for the central systems (Fodor, 2000 gives a book-length treatment of this issue).

### 3.3 Debates about modularity

Fodor’s account of modularity has been very influential, but also very controversial. The criticisms can, I think, be usefully divided into: (1) those that deny that modular input systems are informationally encapsulated; (2) those that reject the importance Fodor accords to informational encapsulation in the characterization of modular systems; and (3) those that deny that the central systems are non-modular. I’ll discuss (1) and (2) but leave (3) to one side except for a brief comment.

#### 3.3.1 Arguments against informational encapsulation

Marslen-Wilson and Tyler (1987) review evidence from a number of word-monitoring experiments (where participants have to monitor speech for a particular

---

<sup>4</sup> Our theory of mind is supposed to be what makes possible our everyday predictions and explanations of one another’s behaviour, such as explaining my journey to the fridge by saying that I desire beer and believe there is beer in the fridge (see Davies and Stone, 1995).

word, and press a button as soon as they hear it) and argue they show pragmatic inferences occurring very early in the processing of the incoming speech stream – i.e. that pragmatic processing is fast. The basic finding of such experiments is that setting up an appropriate prior discourse context will affect speed of recognition of a target word. They report, for instance, that the target word LEAD is detected faster when it occurs in a sentence such as ‘The lead was stripped off the roof’ than in a semantically anomalous sentence such as ‘No lead puzzles some in the land of the text’; but this happens only when the sentence is preceded by an appropriate discourse context such as ‘The church was broken into last night’. They also report that the target word GUITAR is recognized faster in the sentence ‘The young man carried the guitar’ than in the sentence ‘The young man buried the guitar’. The only difference between these latter two sentences is in the relationship between the verb and the target word. Whilst it is quite normal to carry a guitar, it would be somewhat unusual to bury a guitar. The sentences do not differ in either syntactic or semantic appropriateness.

Why are these results relevant to Fodor’s informational encapsulation hypothesis? On Fodor’s picture, the language module outputs, roughly, a representation of the literal meaning of a sentence. This means that the language module does not integrate a sentence with the discourse context – that is the job of the central systems. But if discourse context is processed by the central systems, then information about discourse context should not influence the processing of the language-input module, since the latter is encapsulated from the former. Yet Marslen-Wilson and Tyler’s results seem to show that discourse context does affect the operation of the language-input system, in that it affects reaction times to target words in their monitoring experiments. The idea with regard to the first example above (with the target word LEAD) is presumably, and intuitively, that if you’ve just read a sentence about churches being broken into, then a discourse context – that churches are often broken into and when they are they often have their lead roofs stolen – is constructed, leading to lower RTs to the target word.

So, it seems that Fodor either has to accept that language perception is unencapsulated or say that discourse context is computed by the language-input module. It seems to be Hobson’s choice for Fodor. He *has* to deny the first alternative and find a way to motivate the second. Can he do this? Might it be, as with the examples of context effects on word recognition discussed in Section 3.1.2, that facilitation of recognition is due to the operation of semantic links between words in the mental lexicon? Is this plausible? It’s asking us to believe that the mental lexicon has associative links between two nouns (‘churches’ and ‘lead’), and a verb (‘stolen’). This does not seem immediately appealing. Moreover, Marslen-Wilson and Tyler also present evidence that the resolution of the reference of anaphoric words (words like ‘he’ and ‘she’ that refer back to someone mentioned earlier) is also influenced by discourse context – and it seems extremely unlikely that these words would have associative links in the mental lexicon.

Another response Fodor might make relies not on showing that context effects arise within the language module but that they arise outside the module. In effect this would imply that context effects are not ‘distinctively perceptual’, but ‘post-perceptual’ (Fodor, 1990, p.204). Perhaps the difference in the speed of response to

GUITAR in plausible as opposed to implausible sentences is due to ‘the hearer’s inability to believe that the speaker could have said what it sounded like he said’ (*ibid.*, p.204).

These responses to putative counterinstances to informational encapsulation rely on it being possible to draw a clear, non-arbitrary boundary around the operations of a module.

### ACTIVITY 17.5

---

The Marslen-Wilson and Tyler data concern the language-input module. Can you find any evidence – from this book or elsewhere – that is relevant to Fodor’s claim that the *visual*-input system is informationally encapsulated?

#### COMMENT

You might start by looking again at Chapter 3, Section 5 on ‘Constructivist approaches to perception’. There appears to be evidence there that knowledge affects what we perceive. How might Fodor respond?

---

### 3.3.2 Domain specificity not informational encapsulation

If modular input systems turn out not to be informationally encapsulated, then that’s damaging to Fodor’s position on modularity; but it might not be curtains for modularity itself. One strategy for a supporter of modularity, who is persuaded by the kind of data provided by Marslen-Wilson and Tyler, is to deny the importance of informational encapsulation for modularity. This is the route taken by Max Coltheart who proposes that we should define ‘module’ ‘as a cognitive system whose application is domain-specific’ (Coltheart, 1999, p.118). As Coltheart makes clear, this is to be understood as the very bold conjecture that domain specificity is a *necessary condition* of modularity.

If this account is going to buy us anything, then we need a good account of domain specificity – one grounded, if possible, in the data rather than solely in intuitions about which input stimuli can be classed together as a domain. Coltheart believes that we can get such an account. We do, indeed, start off with educated intuitions about candidate domains. For instance, we might begin with the intuition that visual recognition is a module because visual stimuli constitute a specific domain. We hypothesize, that is, that the visual-input module processes objects, faces and printed words. But then, in the light of evidence from neuropsychological case studies presenting patterns of selective impairment, we might fractionate the visual system into smaller modules. For instance, we might find case studies where a patient has impaired object recognition but intact word recognition and face recognition; another patient who has impaired face recognition but intact object recognition and word recognition; and then a third who has impaired word recognition and intact object and face recognition. This pattern of impairments suggests (if one is convinced of the power of evidence from double dissociations) that there are three domain-specific modules, one for face stimuli, one for objects and one for words.

Coltheart's proposal has the virtues of boldness and simplicity, and it also conforms more exactly to the use cognitive psychologists make of the term 'module'. It is, therefore, also a conservative thesis fitting in with existing bodies of psychological doctrine. We might say that Coltheart's proposal has the Fodorian virtues of being isotropic and Quinean!

### 3.3.3 Prospects for the psychology of the central systems

As we have seen, Fodor believes that the central systems are non-modular. The truth of this is surely moot, given the evidence I mentioned at the end of Section 3.2 that appears to show that there can be specific damage to domain-specific central systems. Moreover, Fodor's arguments for the non-modularity of the central systems rely heavily on the point that information has to be integrated, and the common-sense claim that this requires non-modular systems. But recently Dan Sperber has argued that it is possible to envisage how information might be integrated, at least partially, by central modules if we conceive of these central modules as domain-specific bodies of knowledge that take input from more than one perceptual-input system. And he points out, much as I did at the end of Section 3.2, that this thesis:

gets support from a wealth of recent work ... tending to show that many basic conceptual thought processes found in every culture and in every fully developed human are governed by domain-specific competences. For instance it is argued that people's ordinary understanding of an inert solid object, of the appearance of an organism, or of the actions of a person are based on three distinct mechanisms: a naive physics, a naive biology, and a naive psychology.

*(Sperber, 1996, p.123)*

It's worth noting that these proposed modules are often considered to be domain-specific bodies of knowledge, rather than processes. Whether this gives Fodor any room for manoeuvre, since he makes clear that modules in his sense are *processing* modules and not bodies of knowledge, is a good discussion point. Perhaps a principled distinction can be convincingly drawn between central processing mechanisms (mechanisms for belief fixation, for example) that are non-modular and domain-specific, modular bodies of knowledge.

## Summary of Section 3

---

- Fodor proposes that the mind can be divided into sensory transducers, modular input systems and non-modular central systems.
- Modular input systems are characterized by a cluster of properties, but the properties of being domain specific and informationally encapsulated are especially important.
- Informational encapsulation means that the processing undertaken by a module cannot be affected by knowledge stored in the central systems or by

processing in other modules. Top-down information flow within a module is permitted.

- The central systems receive output from the modular input systems and are involved in belief formation and decision making. They are domain general and unencapsulated. Fodor thinks that a good model for their functioning is the kind of non-demonstrative inference that is used in scientific confirmation. This kind of inference has the properties of being isotropic and Quinean. Fodor's first law of the non-existence of cognitive science warns that the prospects for a computational psychology of the central systems are bleak.
- Critics of Fodor's account of modularity have questioned whether input systems are really informationally encapsulated and suggested that modularity is better defined just in terms of the notion of domain specificity.
- Fodor's picture of the central systems as non-modular has been questioned on the basis that there is evidence for domain-specific bodies of knowledge and on the basis of neuropsychological evidence of the selective impairment of putative central systems.

## 4 Cognitive psychology and the brain

### 4.1 Levels of explanation

In this section I want to discuss the relationship between cognitive psychology and the study of the brain – i.e. neuroanatomy or the study of neural circuits (I'll call this neurobiology in what follows). If you look back over the previous chapters you will find some, though relatively little, discussion of the way in which cognitive systems are actually implemented by the brain. Why is this? There are, I think, two related reasons. The first is the influence of David Marr's meta-theoretical views on how psychologists should investigate information-processing systems. As you saw in Chapter 1, Marr proposed that an information-processing system can be understood at any one of three levels of description. These are:

- **Level 1:** The level of computation where one asks what a device does and why.
- **Level 2:** The level of representation and algorithm where one asks how the computations described in level 1 are implemented; specifically, one is interested in what are the representations of the input and the output to the device, and what the algorithm is for the transformation of those representations.
- **Level 3:** The level of hardware implementation – in psychology this is the level where we would describe the physical realization in the brain of the representations and algorithms described at level 2.

Marr saw important relationships between these levels. In particular, he was clear that '[s]ome types of algorithm will suit some physical substrates better than others'

(Marr, 2003, p.113; first published 1982). This suggests that if one is concerned with a correct theory of human information processing, then level 3 can, in principle, constrain levels 1 and 2 – information about the organization and operation of the brain can constrain what is said at the other levels. But Marr also thought that levels 1 and 2 took precedence over level 3 since, without descriptions at those levels, ‘there can be no real understanding of the function of all those neurons’ (Marr, 2003, p.111). Indeed, Marr took the view that work at level 3 is guided by work at levels 1 and 2 – work at levels 1 and 2 set the agenda, so to speak.

The second reason for the absence of discussion of the organization and function of the physical brain in cognitive psychology is that, until very recently, the techniques and tools were unavailable to undertake investigations at level 3 that would speak to issues concerning human cognitive functioning. In essence, experimental investigations were limited to animal studies where invasive experimentation such as lesion studies and single-cell recording of neurons could take place. These kinds of studies could not, of course, be carried out on humans.

However, the situation has changed over recent years with the advent of neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) that allow images of the human brain to be produced during performance of cognitive tasks. In conjunction with this, there has been a growth in interest in more neurally plausible cognitive models, such as connectionist models. So the question of the relationship between the kinds of cognitive models discussed in this book and models that are pitched at Marr’s level 3 has become a matter of significant theoretical dispute.

## 4.2 The co-evolution of cognitive and neurobiological theories

Although the general question of the relationship of the mind and the brain is one fraught with philosophical difficulties, it is commonly assumed by theorists in the cognitive sciences that cognitive functioning is realized in neural functioning. So Marr is surely right that, in principle, things we find out about neural functioning might constrain cognitive modelling.

For example, consider the phenomenon of covert recognition of familiar faces in some individuals with prosopagnosia (i.e. people who can no longer overtly identify once familiar faces; see Chapter 4, Section 6). Bauer (1984) found that some of his patients presented a heightened autonomic response for once familiar faces compared with unfamiliar faces. Other psychologists (Ellis and Lewis, 2001 provide a succinct review) have reported cases that provide evidence for the covert recognition of once familiar faces on a range of behavioural measures; for instance, better learning of correct as opposed to incorrect face–name pairs. This pattern of evidence might be interpreted in terms of there being two separable information-processing routes responsible for face recognition – one route that generates conscious or overt face recognition and the other that results in unconscious or covert recognition, perhaps because it responds to the emotional significance of a face (Ellis and Lewis, 2001). Bauer proposed that two separate neuroanatomical

pathways – a so-called ventral route involving connections between visual cortex and the limbic system and a so-called dorsal route via the inferior parietal lobule – implement these two cognitive routes.

This example shows cognitive psychology and the study of the organization of neural hardware working in tandem. Clearly, it would be a criticism of Bauer's theory if separate neural pathways were not found, because we would have no explanation of how there could be two separate processing streams.<sup>5</sup> Thus, level 3 constrains cognitive theories pitched at levels 1 and 2. But the example also shows the way in which the study of the neural hardware is *guided* by cognitive modelling. It is hard to see how an investigation at the level of neural hardware could ever get started without there first being a cognitive model (however sketchy) of cognitive functioning.

This view of the relative priority of cognitive psychology over neurobiology accords with that of Coltheart and Langdon when they say that 'it can be very hard to understand what a system is actually doing if one's only information about it is a description of the physical-instantiation level' (Coltheart and Langdon, 1998, p.150) and with Patricia Churchland's comment that 'neuroscience needs psychology because it needs to know what the system does' (Churchland, 1986, p.373).

It is important to emphasize that this view, which I will call (following Patricia Churchland, 1986) the co-evolution of theories view, advocates reciprocal interaction between cognitive psychology and neurobiology. The relative priority accorded to cognitive psychology above is not in the realm of the justification of theory, but in theory discovery. But even here, as neurobiological theory and data accumulate, one can expect neurobiology to play an increasingly important role at the very earliest stages of theory construction. Neurobiologists should not take away the idea that theory and evidence pitched at the cognitive and behavioural level has any kind of evidential or justificatory priority over neurobiology – on this view it doesn't. Stone and Davies summarize this view as follows: 'Cognitive psychology is constrained by neurobiology because neurobiology tells us about the mechanisms in virtue of which psychological generalizations are true. In practice this is constraint without government; challenges and insights flow in both directions' (Stone and Davies, 1999, p.850).

However, this is not to say that the interaction between neurobiological data and cognitive theory is easy or straightforward. A simple example will illustrate this. As we saw earlier in this chapter, Jerry Fodor argues that perception is informationally encapsulated from our beliefs and knowledge about the world. In one of a number of responses to this claim, the neurophilosopher Paul Churchland suggested that, in addition to neuronal pathways that 'ascend' from the retina to primary visual cortex (via structures such as the lateral geniculate nucleus), there is evidence from studies of neuronal cell-staining that there are also matched 'descending pathways' that 'lead us stepwise back through the intermediate brain areas and all the way to the

---

<sup>5</sup> Ian Gold (personal communication) pointed out to me that this would not be decisive. It is perfectly possible for a single neural pathway to have neurons that have different firing modes depending on different neurotransmitters, for example. Thus, a single route might implement two different cognitive functions via different modes of firing.

earliest processing systems at the retina' (Churchland, 1989, p.266). Churchland then interprets these 'descending pathways' as feedback pathways from the central systems to the input systems that 'strongly suggest' (*ibid.*) that perceptual processes are not encapsulated.

Fodor's reaction to this neurobiological objection to informational encapsulation is instructive. He says: 'Heaven knows what psychological function "descending pathways" subserve. ... One thing is clear: if there is no cognitive penetration of perception then at least "descending pathways" aren't for that' (Fodor, 1990, p.261).

One way to interpret this is to see Fodor as denying, in principle, the relevance of neurobiological data for cognitive psychology. But this is unlikely to be correct given that Fodor is famous for his view that it is not possible 'to enumerate a priori the kinds of facts a scientific theory is required to account for' (1985b, p.147; first published 1981). (Recall also the isotropy of scientific confirmation mentioned in Section 3.2.) What Fodor is suggesting – in accord with the co-evolution view – is that *as matters currently stand* our understanding of the role of the descending pathways is too primitive to be taken as evidence against informational encapsulation. Moreover, in accord with the relative priority of cognitive theory over neurobiological data in the order of discovery, identifying the role of those pathways requires knowing what functions they subserve. If we could find independent reasons for thinking that those pathways did carry information from central systems to input systems, then their evidential role would change.

### 4.3 The radical neuron doctrine

The 'co-evolution of theories' view discussed in the previous section sees reciprocal influence between cognitive psychology and neurobiology, but also maintains that there is a limited heuristic priority of cognitive *theory* over neurobiological *data* – at least for now. In contrast to this is a view Gold and Stoljar (1999) dub the 'radical neuron doctrine'. The doctrine is also known as 'eliminative materialism' (see Chapter 15, Section 1.2.2), and is usually associated with the neurophilosopher Paul Churchland (e.g. Churchland, 1989). This is the doctrine that:

a successful theory of the mind will be a theory of the brain expressed *in terms of* the basic structural and functional properties of neurons, ensembles or structures, [that] neurophysiology, neuroanatomy, and neurochemistry will by themselves eventually have the conceptual resources to understand the mind and, as a consequence, a successful theory of the mind will make no reference to anything like the concepts of ... the psychological sciences as we currently understand them.

*(Gold and Stoljar, 1999, p.814)*

This doctrine implies that contemporary cognitive psychology is a historical staging post on the way to the terminus of a neurobiological theory of the mind, and will disappear when that end point is reached. On this view, rather than the co-evolution of theories, we have the extinction of cognitive theories as neurobiological theories gradually take over.

Some theorists seem to think that the radical neuron doctrine follows from a fundamental commitment that most contemporary cognitive scientists share. As I mentioned earlier, just about everyone thinks that mental functioning is the result of the physical organization and functioning of the brain. It can then seem to be obvious that, ultimately, a theory of the mind will be produced that talks only about the physical parts of the brain – neurons, neural circuits, neurotransmitters, etc.

But this would be to move far too fast. To use an example of Gold and Stoljar's: earthquakes are made up of millions of physical particles that behave in accordance with the laws of physics. But the science of earthquakes shows no signs of being replaced by physics. And the reason for this is quite general: just because something is built out of lots of Xs (physical particles), it does not follow that we will get an understanding of that thing in terms of the science of Xs. Indeed, if this were so, then neurobiology would ultimately be replaced by physics, since the physical components of the brain are composed of exactly the same stuff as the rest of the physical universe – protons, electrons, bosons, quarks, or whatever current physics tells us are the fundamental building blocks of the universe.

But the supporter of the radical neuron doctrine has other arguments. I will briefly discuss only one of these. This is the idea that theories in cognitive psychology will *reduce* to theories in neurobiology. The topic of inter-theoretic reduction is a complex one (Churchland and Churchland, 1998, give a basic introduction). But the basic thought is that 'because everything in the world is made of the same basic stuff in complex combinations, the laws of biology ought to be derivable from those of chemistry, and the laws of chemistry from the laws of physics' (Ladyman, 2002, p.95). Similarly, the idea is that the laws of cognitive psychology ought to be logically derivable from theories in neuroscience. Examples from the history of science that are usually adduced to illustrate and motivate this view are the reduction of biology to molecular biology and the reduction of chemistry to quantum mechanics. Notice that this is not the same idea as the one we discussed in the previous paragraph. Whilst it starts from the premise that the world is made up from the same basic kinds of stuff, it gets to the conclusion that psychology should be reduced to neuroscience via the idea that psychological theories can be logically derived from neurobiological theories. It remains the case that the mere fact that Xs and Ys are made up of the same stuff doesn't allow us to draw the conclusion that we can understand Xs in terms of Ys. But problems for this position remain. It is still the case that there seems to be no reason in principle to think that reduction of theories will stop at neurobiology. Could the friend of neurobiological reduction just dig in her heels here and say that it will turn out that whilst cognitive psychology reduces to neurobiology, neurobiological theories are not, as a matter of fact, going to be logically derivable from theories in physics? Well, she could, but there seems no reason for thinking this to be true if you buy into reductionism in the first place.

The linguist Noam Chomsky (2000, 2002) has articulated another major problem for reductionism. He argues that inter-theoretic reduction is historically rare, and that where there have been genuine cases of reduction – such as that of chemistry to physics – this was only possible when there was a radical change in physics. The moral Chomsky draws from this is that reduction is not to be aimed for nor expected. The only sensible aim is for each theoretical enterprise to pursue

its own path (as did chemistry and physics). It cannot be ruled out that there will be what Chomsky calls the ‘unification’ of theories (as with chemistry and physics), but nor can we be at all certain that human intelligence will be up to the task of unification. And so far as psychology and neurobiology are concerned, ‘one can entertain the idea that “the mental is the neurophysiological at a higher level”, but for the present, only as a guide to enquiry, without much confidence about what “the neurophysiological” will prove to be’ (Chomsky, 2003, p.265).

Chomsky’s point can, perhaps, be put in the following way. The reduction of one theory, T<sub>1</sub>, to another theory, T<sub>2</sub>, requires that both T<sub>1</sub> and T<sub>2</sub> are successful theories in their respective domains. If either or both are not successful, then why would one either expect or want reduction? With regard to psychology and neurobiology, our theories are tentative and far from being in the state where a reduction is, even in principle, in the offing. So, as things currently stand, the claim that psychology will reduce to neurobiology is merely a *historical speculation*, and one based on very scant evidence from the history of science.

There is one final twist in the dialectic that we need to mention. I suggested above – when discussing the idea that a neurobiological reductionist may have to accept that neurobiology will in its turn be reduced to physics – that a determined neurobiologist might simply dig in her heels. There is a major strand of thought (composed of many different fibres) in both psychology and the philosophy of mind that suggests that a determined psychologist can dig in her heels and *deny* that psychology is in principle a candidate for reduction to neurobiology. The denial is based on an argument that there is something special about the psychological domain that will forever defeat reduction. One way in which this argument can go stems directly from the CMM. The CMM is committed to the idea that psychological states are *multiply realizable* – that is, that a creature made from entirely different stuff from us (silicon-based stuff and not carbon-based stuff, for example) may have a psychology (at least a *cognitive* psychology) very similar to ours. In other words, a psychology like ours might be found in a multiplicity of different physical embodiments. Recall the point made earlier in the chapter (at the end of Section 2.1) that the CMM abstracts away from concerns about our physical make-up. This suggests that a cognitive psychology that aimed to be truly general would have to model the psychology of creatures regardless of what they were made from – creatures with wholly different physical composition. It follows from this that ‘neurobiology’ would have to describe physical properties that apply to all of these different physical compositions. But this is implausible – how could such a science describe the common physical properties of, say, carbon-based and silicon-based creatures? (Fodor, 1975, gives a seminal statement of this view.) Of course, while multiple realizability seems to undermine the prospects for reducing psychology in general, it doesn’t follow that a psychology of specifically *human* information processing could not, in principle, be reduced to a neurobiology of the *human* brain.

## Summary of Section 4

---

- Marr's views suggest that cognitive models take precedence over neurobiological models of cognitive functioning. But Marr also thought that neurobiological models could constrain cognitive modelling.
- The co-evolution of theory view sees reciprocal interaction between neurobiological and cognitive models, though cognitive psychology may have relative priority, at least in the order of discovery.
- The radical neuron doctrine predicts that cognitive psychology will eventually be replaced by neurobiology.
- There are reasons to doubt this prediction, both on theoretical and historical grounds.

## 5 Conclusion

This chapter has introduced you to a selection of the main theoretical debates that have taken place in cognitive psychology and in cognitive science more broadly. None of these debates has a resolution that carries a consensus amongst cognitive psychologists/scientists. There are many other important theoretical debates that warrant attention. Amongst those that I have most reluctantly omitted are those over the nature of the medium of thought itself (the debate over the so-called language of thought hypothesis), the debate over whether current psychological theories of concepts can meet the compositionality constraint, and very recent debates on whether cognitive psychology will be fundamentally reconfigured if we give greater emphasis to our embodiment. Readers who have developed a taste for theoretical issues can consult Rey (1997), the introduction to Laurence and Margolis (1999), and Clark (2001) for introductions to these debates.

What I hope that you take away from this chapter is a greater insight into the fundamental principles that lie behind theories in contemporary cognitive psychology, and that this will facilitate greater understanding of the individual topics covered in this book.

### Acknowledgement

Many thanks to Ian Albery, Max Coltheart and Ian Gold for comments on parts of initial drafts of this chapter. Special thanks to Martin Davies who kindly commented on the whole of the initial draft.

## Further reading

- Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds) *An Invitation to Cognitive Psychology, Volume 3, Thinking*, Cambridge, MA, MIT Press.
- Coltheart, M. (1999) Modularity and cognition', *Trends in Cognitive Sciences*, vol.3, pp.115–20.
- Marslen-Wilson, W. and Tyler, L.K. (1998) 'Rules, representations, and the English past tense', *Trends in Cognitive Sciences*, vol.2, pp.428–35.

## References

- Baron-Cohen, S., Tager-Flusberg, H. and Cohen, D. (eds) (1999) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (2nd edn), Oxford, Oxford University Press.
- Bauer, R.M. (1984) 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia*, vol.22, pp.457–69.
- Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds) *An Invitation to Cognitive Psychology: Volume 3 – Thinking*, Cambridge, MA, MIT Press.
- Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.
- Chater, N. and Ganis, G. (1991) 'Double dissociation and isolable cognitive processes', *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ, Lawrence Erlbaum.
- Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*, Cambridge, Cambridge University Press.
- Chomsky, N. (2002) *On Nature and Language*, Cambridge, Cambridge University Press.
- Chomsky, N. (2003) 'Replies', in Antony, L.M. and Hornstein, N. (eds) *Chomsky and his Critics*, Oxford, Blackwell.
- Churchland, P.M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press.
- Churchland, P.M. and Churchland, P.S. (1998) 'Inter-theoretic reduction: a neuroscientists field guide', in Churchland, P.M. and Churchland, P.S. *On the Contrary: Critical Essays, 1987–1997*, Cambridge, MA, MIT Press.
- Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA, MIT Press.
- Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford, Oxford University Press.
- Cohen, G., Johnson, R.A. and Plunkett, K. (eds) (2000) *Exploring Cognition: Damaged Brains and Neural Networks: Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press.

- Coltheart, M. (1999) 'Modularity and cognition', *Trends in Cognitive Sciences*, vol.3, pp.115–20.
- Coltheart, M. and Davies, M. (2003) 'Inference and explanation in cognitive neuropsychology', *Cortex*, vol.39, pp.188–91. (*Cortex* is freely available on the Word Wide Web.)
- Coltheart, M. and Langdon, R. (1998) 'Autism, modularity and levels of explanation in cognitive science', *Mind and Language*, vol.13, pp.138–52.
- Coltheart, M., Curtis, B., Atkins, P. and Haller, M. (2000, first published 1993) 'Models of reading aloud: dual-route and parallel-distributed-processing approaches', in Cohen, G., Johnson, R.A. and Plunkett, K. (eds). (First published in *Psychological Review*, vol.100, pp.589–608.)
- Crane, T. (2003) *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations* (2nd edn), London, Routledge.
- Davies, M. (1990a) 'Knowledge of rules in connectionist networks', *Intellectica*, vols 9–10, pp.81–126.
- Davies, M. (1990b) 'Rules and competence in connectionist networks', in Tiles, J.E., McKee, G.T. and Dean, G.C. (eds) *Evolving Knowledge in Natural Science and Artificial Intelligence*, London, Pitman.
- Davies, M. (1995) 'Two notions of implicit rules', in Tomberlin, J. (ed.) *Philosophical Perspectives, Volume 9: AI, Connectionism and Philosophical Psychology*, Oxford, Blackwell.
- Davies, M. and Stone, T. (eds) (1995) *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell.
- Ellis, A.W. and Young, A.W. (1998) *Cognitive Neuropsychology: A Textbook with Readings*, Hove, Psychology Press.
- Ellis, H.D. and Lewis, M.B. (2001) 'Capgras delusion: a window on face recognition', *Trends in Cognitive Sciences*, vol.5, pp.149–56.
- Fodor, J.A. (1975) *The Language of Thought*, Cambridge, MA, Harvard University Press.
- Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MA, MIT Press.
- Fodor, J.A. (1985a) 'Précis of *The Modularity of Mind*', *The Behavioral and Brain Sciences*, vol.8, no.1, pp.1–42. (Reprinted in Fodor, 1990, pp.195–206.)
- Fodor, J.A. (1985b, first published 1981) 'Some notes on what linguistics is about', in Katz, J. (ed.) *The Philosophy of Linguistics*, Oxford, Oxford University Press. (First published in Block, N. (ed.) *Readings in the Philosophy of Psychology*, vol. II, Cambridge, MA, Harvard University Press.)
- Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA, MIT Press.
- Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA, MIT Press.
- Fodor, J.A. (2000) *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*, Cambridge, MA, MIT Press.
- Fodor, J.A. and Pylyshyn, Z.W. (1988) 'Connectionism and cognitive architecture: a critical analysis', *Cognition*, vol.28, pp.3–71.

- Gallistel, C.R. (2001) ‘Mental representations, psychology of’, in Smelser, N.J. and Baltes, P.B. (eds) *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam and New York, Elsevier.
- Garson, J. (2002) ‘Connectionism’, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*, [online] Available from: <http://www.seop.leeds.ac.uk/entries/connectionism/> [Accessed 10 March 2004]
- Gold, I. and Stoljar, D. (1999) ‘A neuron doctrine in the philosophy of neuroscience’, *Behavioral and Brain Sciences*, vol.22, pp.809–69.
- Johnston, R.A. and Braisby, N. (2000) ‘Introduction’, in Cohen, G., Johnson, R.A. and Plunkett, K. (eds).
- Juola, P. and Plunkett, K. (2000) ‘Why double dissociations don’t mean much’, in Cohen, G., Johnson, R.A. and Plunkett, K. (eds).
- Ladyman, J. (2002) *Understanding Philosophy of Science*, London, Routledge.
- Laurence, S. and Margolis, E. (eds) (1999) *Concepts: Core Readings*, Cambridge, MA, MIT Press.
- Macdonald, C. and Macdonald, G. (eds) (1995) *Philosophy of Psychology: Debates on Psychological Explanation*, Oxford, Blackwell.
- Marr, D. (2003, first published 1982) ‘The philosophy and the approach’, in Yantis, S. (ed.) *Visual Perception: Essential Readings*, Hove, Psychology Press. (First published as Chapter 1 of Marr, D., 1982, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company.)
- Marslen-Wilson, W. and Tyler, L.K. (1987) ‘Against modularity’, in Garfield, J. (ed.) *Modularity in Knowledge Representation and Natural Language Understanding*, Cambridge, MA, MIT Press.
- Marslen-Wilson, W. and Tyler, L.K. (1998) ‘Rules, representations, and the English past tense’, *Trends in Cognitive Sciences*, vol.2, pp.428–35.
- McClelland, J.L. and Patterson, K. (2002a) ““Words or rules” cannot exploit the regularity of exceptions”, *Trends in Cognitive Sciences*, vol.6, pp.464–5.
- McClelland, J.L. and Patterson, K. (2002b) ‘Rules or connections in past-tense inflections: what does the evidence rule out?’, *Trends in Cognitive Sciences*, vol.6, pp.465–72.
- Pinker, S. and Prince, A. (1988) ‘On language and connectionism; analysis of a parallel distributed processing model of language acquisition’, *Cognition*, vol.28, pp.73–193.
- Pinker, S. and Ullman, M.T. (2002a) ‘The past and future of the past tense’, *Trends in Cognitive Sciences*, vol.6, pp.456–63.
- Pinker, S. and Ullman, M.T. (2002b) ‘Combination and structure, not gradedness, is the issue’, *Trends in Cognitive Sciences*, vol.6, pp.463–74.
- Plunkett, K. and Marchman, V. (1993) ‘From rote learning to system building: acquiring verb morphology in children with connectionist nets’, *Cognition*, vol.48, pp.21–69.
- Rey, G. (1997) *Contemporary Philosophy of Mind*, Oxford, Blackwell.

- Rumelhart, D.E. and McClelland, J.L. (1986) ‘On learning the past tenses of English verbs’, in Rumelhart, D.E. and McClelland , J.A. (eds) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA, MIT Press.
- Searle, J.R. (1992) *The Rediscovery of the Mind*, Cambridge, MA, MIT Press.
- Seidenberg, M.S. and MacDonald, M.C. (1999) ‘A probabilistic constraints approach to language acquisition and processing’, *Cognitive Science*, vol.23, pp.569–88.
- Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press.
- Sperber, D. (1996) *Explaining Culture: A Naturalisitic Approach*, Oxford, Blackwell.
- Stone, T. and Davies, M. (1999) ‘Autonomous psychology and the moderate neuron doctrine’, *Behavioral and Brain Sciences*, vol.22, pp.849–50.
- Swinney, D.A. (1979) ‘Lexical access during sentence comprehension: (re)consideration of context effects’, *Journal of Verbal Learning and Verbal Behavior*, vol.18, no.5, pp.645–59.
- Ullman, M.T., Corkin, S. Coppola, M., Hickol, G., Growdon, J.H., Koroshetz, W.J. and Pinker, S. (1997) ‘A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system’, *Journal of Cognitive Neuroscience*, vol.9, pp.266–76.



# Epilogue

In Chapter 1 of this volume, we attempted to lay out some of the foundations of cognitive psychology, in part by tracing some of its historical antecedents. We hope that, having reached the end of the book, you now find several of the features we described in that first chapter are in much sharper focus.

One fundamental aspect of mind is its intentional nature – mental states are *about* aspects of the world and the discussions throughout this volume have illustrated this. In Part 1, we saw how the visual system recovers information about perceived objects from the patterns of light that fall on the retina. Part 2 outlined the processes by which the language system recovers information about perceived words from patterns of sound (in audition) or light (in vision). Part 3 was concerned with our memories of particular aspects of (external) events. Even Part 4, concerned with the seemingly inwardly directed activity of thinking, was similarly concerned with the external aspect of mental representations: thinking about problems, choices and arguments involves recovering true information about such things in the form of correct solutions, likelihoods, and valid conclusions.

Another aspect of mind that has become clear throughout this volume is its fractionation. In part, the strategy of isolating particular mental faculties from one another is a methodological one, pursued by cognitive psychologists in order to facilitate systematic study. But cognitive psychologists also believe that the mind *is* fractionated, containing multiple interacting components that, acting in concert, give rise to what appears to us to be a unified mind. The different parts of this book also reflect the fractionation – we consider visual perception in isolation from language comprehension partly for reasons of methodological convenience, but partly because we believe that visual perception and language comprehension call on unique cognitive processes.

The fractionation of the mind has been elevated to the status of a philosophical thesis – the so-called modularity of mind – that was the subject of explicit discussion in Chapter 17, and much implicit discussion elsewhere. The modularity thesis distinguishes input systems, such as visual perception and the early stages of language comprehension, from central systems, such as reasoning, judgment and arguably certain aspects of categorization. The extent that mental faculties reflect distinct processing modules adds further justification to studying them in isolation from one another.

If the mind can be characterized in terms of modularity, the question arises as to what kind of modules it possesses. Most of the chapters in this book have shown that certain cognitive processes are subject to top-down influence, and the influence of quite general knowledge. While such influences do not necessarily undermine a strict interpretation of modularity, one in which processes are informationally encapsulated, they nevertheless suggest that alternative understandings of modularity should be considered. Perhaps there is a sense of modularity in which modules can influence one another, or can influence one another to a certain degree. The exact nature of the mind's modularity remains an open question – that it has some kind of modular structure seems to be a claim on which many different theorists agree.

In Chapter 17 we also saw an argument that our understanding of central systems is likely always to be poor. If central systems such as reasoning are sufficiently flexible as to be able to call on any aspect of general knowledge, if they are so informationally *un-encapsulated*, the argument goes, then it may not be possible to delineate them in a precise or meaningful way. However, much of this book suggests the opposite. Where we have seen the influences of general knowledge on systems such as reasoning, the result has often been an improved understanding of the processes involved. Even straying from the prime candidates for modularity, perceptual systems and the early stages of language comprehension, researchers have succeeded in identifying systematic influences on processing. In reasoning, for example, researchers have developed theories to account for conditional inference, and have succeeded in devising tasks that evaluate these. The evaluation may be incomplete, but the picture their research suggests is very different from the negative one offered by the modularity thesis.

The chapters in this book have also illuminated the preferred kinds of explanation of cognitive processes. Chapter 1 outlined the importance of building models in the development of scientific theories and, throughout the book, more or less detailed models of different cognitive processes have been considered. Generally, researchers have developed two kinds of computational model for cognitive processes – symbolic and connectionist. These offer very different kinds of understanding of cognitive processes, and two different means of explaining them. In symbolic models, cognition involves the rule-based manipulation of (neurally realized) symbols and symbol expressions. In connectionism, cognition involves the formation of a stable pattern of activation across processing units that, in general, are not symbolic.

As this volume has revealed, there is much debate about the merits of these two styles of computational modelling. Each has arguments in its favour. Symbolic models have advantages in modelling rule-based systems, as language appears to be (though this too is debated). Connectionist models have the advantage of modelling cognition in terms of processes and units that are superficially similar to processes and structures in the brain.

The debate over symbolic and connectionist models hints at a broader tension concerning the kind of explanation at play in cognitive psychology. In Chapter 1, we outlined Marr's three levels of explanation, suggesting that cognitive psychology was concerned more with levels 1 and 2, the computational and algorithmic levels, than with level 3, the hardware or implementation level. Throughout this book, however, we have seen evidence of the close relation between cognitive processes and underlying physiological processes. Indeed, some have questioned whether these are more closely linked than Marr's analysis suggests. For example, connectionists try to ensure that the processes involved in computational models are consistent with what we know of neural processes.

Evidence from people who have experienced brain damage provides valuable evidence of the nature of cognitive processes. If cognition is truly fractionated as researchers believe, then it ought to be possible for certain mental faculties to be impaired through damage whilst leaving others intact. Such dissociations have been reliably observed in many different areas of cognition, and have helped researchers to build, evaluate and refine ever more complex models of cognitive processes.

By providing even more direct evidence of the nature of neural activity, neuroimaging techniques have enriched our understanding of cognitive processing. Whilst participants perform tasks, images can be generated that indicate regions of the brain with greatest neural activity. If researchers have a prior understanding of the function of those brain areas, they can use this evidence to infer what kinds of information and information processing are involved in different tasks. Of course, it is difficult indeed to understand precisely the functions of particular brain areas, and so the interpretation of neuroimaging evidence is hotly debated.

These are just some of the main themes to have emerged from the chapters in this volume. At heart, they are all deeply concerned with the fundamental nature of cognition. How these debates ultimately turn out is likely to have profound implications for our understanding of the mind.

Though these debates are ongoing, the chapters in this book document just how much progress researchers have made in understanding the structure and time course of cognitive processes. It is perhaps this combination of profoundly significant debates with considerable progress on the detail of cognitive processes that makes cognitive psychology such an exciting subject. We hope the chapters of this book have given you a rich understanding of cognitive psychology and, above all, have enabled you to share this sense of excitement and enthusiasm.



# Index

- aboutness **23**  
  intentional nature of mind **655**
- abstract conditional inference task **427–31**  
  and mental logic **428–9**, **439**  
  and mental models **429–30**  
  and the probabilistic approach **431**
- abstract selection task **437–42**  
  and confirmation bias **438**  
  and the matching effect **438**  
  and mental logic **438–9**  
  and mental models **439–40**  
  and the probabilistic approach **440–2**
- AC (affirming the consequent) **423**, **424**  
  and the abstract conditional inference task **427**, **428**, **429**, **430**, **431**  
  and the abstract selection task **438**, **439**, **440**  
  suppression **432**, **433**, **434–5**, **436**
- access consciousness **546**, **547–8**
- ACME (analogical constraint mapping engine) **365**
- ACT theory of skill acquisition **372**, **375**
- ACT-R architecture **460**, **522**, **584**, **585–610**  
  and architectural assumptions **600**  
  and auxiliary assumptions **600**  
  current goal **586**, **587**  
  evaluating **611–13**  
  goal stack **586**, **587**, **591**  
  history of **585**  
  learning and using arithmetic skills **601–8**  
  and list memory **592–9**, **600**  
  and PDP **608–10**  
  and procedural memory **586**, **588–90**  
  production compilation **586–7**  
  running the model **599–600**  
  *see also* declarative memory
- action  
  attention as directing actions **65–6**  
  and perception **85**, **97**, **104**, **105**  
    Gibson's theory of **80–1**, **118**  
    object recognition by touch **118–20**  
    and production rules **583**
- activation equations **597**
- activation level  
  chunks in ACT-R architecture **587–8**  
  and spoken word recognition **204**
- activation threshold, and ACT-R models **598–9**
- adaptive coherence **537–8**
- addition by counting, ACT-R model of **604–7**
- addressed phonology **208**
- adversary problem solving **365**
- affect grid **474–6**
- affective priming **565–6**
- affective tone **458**, **459**
- affordances, and Gibson's theory of perception **89–90**, **104–5**, **113**
- agentic personality types **524**
- AI *see* Artificial Intelligence
- aircraft pilots  
  and attention **42**, **46**, **64**  
  and Gibson's theory of perception **82**
- algorithmic level of cognition **27–9**  
  and Marr's theory of perception **91**, **95**, **96**
- all or nothing behaviour **433–4**
- Allais paradox **390–1**, **398**
- Allport, D.A. **65–6**
- Altmann, G.T.M. **222**
- Alzheimer's disease **327**
- ambient optic array **82–9**  
  flow in the **86–9**  
  and invariant information **82–6**
- ambiguous words  
  and concepts **164**  
  and dialogue **256**  
  and parsing **225**  
  semantic ambiguity **217–18**
- amnesia  
  childhood amnesia **511–12**, **513**  
  and declarative and procedural memory **282–3**  
  and episodic and semantic memory **280–1**  
  and errorless learning **567–8**  
  and implicit memory **290–1**, **554**, **555**  
  and PTSD (post-traumatic stress disorder) **531**, **534**
- anaesthesia, learning in patients receiving **559–60**
- analogical constraint mapping engine (ACME) **365**
- analogical mapping **364–5**
- analogical problem solving **363–5**, **376**, **579**
- anaphora **232**  
  resolution **233–4**
- anchor and adjust heuristic **406**

- Anderson, A. 254  
 Anderson, J.R. 279, 280, 372  
   and ACT-R architecture 585, 591, 592, 600, 609, 610  
   and the Newell Test **612–13**  
 Andrade, J. 323  
 animals  
   and categorization behaviour 165  
   and emotion 479  
   lesioned animals and emotion **501–2**  
 Anolli, L. 363–4  
 anomalies, in language processing 241–2  
 ANS (autonomic nervous system) **465–6**  
 antecedents, and deductive reasoning **421**  
 anti-aircraft gunnery, and cybernetics 15  
 anxiety, and the attentional bias 490–1  
 appraisal theories, and emotion 496–9  
 architectural assumptions **600**  
 arguments, logically valid **421–3**  
 Aristotle 169, 418, 420  
 arithmetic skills, and the ACT-R model 60–18  
 Armstrong, S.L. 178–9  
 arousal **474**  
 articulatory rehearsal loop **314**, 317  
 articulatory suppression **317–19**, 331  
 artificial grammar, and implicit learning **557–9**  
 Artificial Intelligence (AI) 17, 549  
   and the frame problem **638–9**  
 assembled phonology **208**  
 associated words **216**  
 association strength **216**  
   and ACT-R models **598**  
 Atherton, M. 73  
 Atkinson, R.M. 270, 309, 560  
 attention 6, 7, 31, 34, 35, 37–70  
   auditory 34, 37–45  
   and competition 65  
   and consciousness 66–7, 459, 550, 551, 564–5  
   and distraction 59–62  
   and emotion 488–91  
   and Fodor’s theory of modularity 636  
   neurology of 62–5  
   and stored knowledge 35  
   visual 45–59  
   and working memory 325–6  
     fractionation 326–7  
 attentional blink **52**  
 attenuation process **43**  
 attribute-listing, and concepts **166**, 171  
 attribution theory **492–3**  
 audience design, in language production **249–51**, 252  
 auditory attention 34, 37–45  
   attending to sounds 41–2  
   dichotic listening 41, 43, 44, 54, 60, 61  
   disentangling sounds 37–41  
   eavesdropping on the unattended message 43–4  
   and seeing 45–6  
 autobiographical memory 459, 507–43, 618  
   and autobiographical knowledge 517–20, 521  
   cue word experiment 514–17  
   defining 507–9  
   and direct retrieval **525**, **527**  
   and episodic memory 512, 518, 520–2, 526, 528–9, 537–8  
   flashbulb memories 507–8, 531  
   and Fodor’s theory of modularity 636  
   and generative retrieval **525–7**, 528, 529  
   and the lifespan retrieval curve 509–10, **511–14**  
   ‘probe’ experiments 528  
   self-defining memories 508  
   and semantic memory 520–2  
   *see also* PTSD (post-traumatic stress disorder); working self  
 autonomic nervous system (ANS) **465–6**  
 autonomous view, of semantic ambiguity **217–18**  
 auxiliary assumptions **600**  
 availability heuristic **405–6**  
 avoidance symptoms, in PTSD (post-traumatic stress disorder) **533–4**  
 Baars, B.J. 545, 571, 573  
 babies  
   and face recognition 148–9  
   and speech segmentation 202  
 backward masking **47**, 49, 50  
   and perception 106–8  
 backward recall, and ACT-R models **592–3**, 600  
 Baddeley, A.D. 555, 565  
   on errorless learning 567–8  
   on working memory 307, 308, 310–11, 312, 336  
     and consciousness 571, 572, 573

- executive processes 323, 325, 326, 327, 328  
 phonological 317, 318, 319, 320, 321, 329, 330  
 tripartite model of 314–15, 316, 317, 328
- Bahrick, H.P. 136
- Baker, L. 241
- Balint's syndrome 64
- Banks, W.P. 551
- Bard, E.G. 257
- Bard, Philip, and the Cannon-Bard theory of emotion 494–6
- Bargh, J.A. 555
- Barsalou, L.W. 165, 171
- Bartlett, J.C. 151
- base rates **404–5**
- base-level activation of a chunk **597**
- base-rate neglect 402, 406–7
- basic emotions **469–73**, 474, 502
- Bauer, R.M. 144–5, 644–5
- Bayes' Theorem 345, 346, 400–2, 404, 411, 431
- Baylis, G.C. 64
- behavioural control, and consciousness 564–8
- behavioural decision research 392–6
- behaviourism 10–12
- cognitive behaviourists 13–14
  - and complex actions 12–13
  - and consciousness 10, 12, 550, 551
  - and decision making 382
  - and emotion 463
  - and language 12–13
  - science and the unobservable 10–11
  - and TOTE units 15–16
- belief updating, and the frame problem **638–9**
- Berko, J. 214
- between-category distinctions, in object recognition **116**
- Bever, T.G. 225
- bi-conditionals
- and mental logic **428–9**
  - and mental models 429, 430, 440
  - probabilistic approach to 431
- Biederman, I., theory of object recognition 131–4
- binding problem **327–8**, 571–2
- binocular rivalry 571
- Binstead, G. 105
- Bisiach, E. 63
- Bjork, R.A. 286
- Blanchette, I. 364
- Blaxton, T.A. 289
- blind children, facial expressions 472
- blindsight 551, 552, **563**, 568–9
- Block, N. 547–8
- Bluck, S. 519
- body language 472
- Bohr, N. 637
- book bag paradigm **400–2**
- Boole, G. 418, 419, 420
- borderline cases, categorization of 172–3, 181, 190
- bottom-up processing
- and perception **76**, 77, 98
  - Gibson's theory of 80–90
  - Marr's theory of 90–8
  - and visual masking 106–8
  - and visual illusions 98–9
- bottom-up support, and spoken word recognition 205
- bounded rationality **411**, 413
- Bower, G.H. 483, 484, 486, 487, 488, 489, 585
- box (candle) problem 355–6
- the brain
- and attention 62–5
  - and cognitive psychology 460–1, 643–9
  - and the computational model of the mind 620–1
  - and neurobiology 648
  - and perception **76**, 102–8
  - primary visual cortex 102
  - see also* dorsal stream; ventral stream
- brain damage
- and attention 65
  - sensory neglect **62–4**
  - visual displays 56
- and consciousness 546, 562–3, 572
- and the dorsal and ventral streams 103–4
- and face recognition 144–8, 618–19
- and implicit learning 560
- lesioned animals and emotion **501–2**
- and the lexical decision task 627–8
- and memory 269, 299, 310, 325
- and normal cognition 2–5
- and perceptual errors 73
- and perceptual processes 34–5
- see also* amnesia

- brain-imaging techniques 4, 20–2  
 and consciousness 545  
 and emotion 472–3, 503  
 hypnosis and attention 66  
 and memory 267, 280–1  
 and perception 90  
 and the phonological loop 321, 322  
 and reasoning 340  
 and thinking 346  
*see also* fMRI (functional magnetic resonance imaging)
- Braisby, N.R. 185–6
- Brandimonte, M.A. 316
- Branigan, H.P. 257
- Bredart, S. 242
- Brentano, Franz 23, 164
- bridging inferences 236
- Broadbent, D.E. 41, 42, 43, 44, 50, 58
- Broadbent, M.H.P. 50
- Brown, G. 233
- Brown, G.D. 333
- Bruce, V. 90, 114–15, 135, 137  
*Bruce and Young model of face recognition*  
 139–41, 617, 618, 619
- Bruner, J.S. 164
- Brunswik, E. 409
- Buchanan, M. 317
- Budescu, D. 410
- Bullemer, P. 558
- Bulthoff, H.H. 134
- Burgess, N. 333–4
- Burton, A.M. 137
- Bushyhead, J.B. 409
- Byrne, R.M.J. 432–3, 434, 435
- Cacioppo, John 496
- Cannon, William, and the Cannon-Bard theory of emotion 494–6
- canonical coordinate frames 124
- Capgras delusion/syndrome 73, 145–7, 479–80
- Carey, S. 149–51
- Carlton, L.G. 105
- Carpenter, P.A. 310, 311–13, 323–4
- catalogue of 3D models 129–31
- categories, and concepts 163–4
- categorization 1, 6, 34, 158–9, 160, 164–6  
 and categorizers 190  
 concepts and categorization behaviour 164–6  
 and consciousness 459
- in development 183  
 and diagnosis 167  
 different kinds of 188–9  
 and expert problem solving 368–9  
 explaining 169–87  
 and language 158, 165, 189  
 linguistic view of 158  
 perceptual view of 158  
 and recognition 114, 115  
 rule-based mode of 189  
 similarity-based mode of 189  
 and symbol systems 25  
*see also* concepts
- categorization judgements 172–4
- category verification 171
- category-specific impairments 3–5  
 and cognitive modelling 4–5  
 experimental and neuroimaging methods 4  
 neuropsychological methods 3
- central executive, and working memory 314, 323–8, 336, 572
- central (interruption) masking 50
- Cermak, L.S. 291
- chaining hypothesis, and the phonological loop 332–3
- Chalfonte, B.L. 280
- Chalmers, D. 548
- Charness, N. 375
- Chase, W.G. 366, 367, 369
- Chater, N. 445
- Cheesman, J. 47–8
- chemistry  
 and representation 23  
 and the unobservable 11, 20
- Cheshire cat illusion 564, 570–1
- chess  
 and the algorithmic level of cognition 28  
 studies of expert problem solving 365, 366–7, 369–70, 371–2, 376–7
- Chi, M.T. 368, 369, 373
- children  
 and autobiographical memories 510–11  
 childhood amnesia 511–12, 513  
 and categorization 183, 185, 187  
 facial expressions 472  
 learning addition by counting 604–7  
 learning past tense verbs in English 607–8, 608–9  
 and second language learning 330

- and visuo-spatial working memory 320  
and word-length effects 322–3
- Chomsky, Noam 13, 17, 221, 222, 647–8
- Christensen-Szalanski, J.J.J. 409
- Chronicle, E.P. 361–2
- chunks in ACT-R models **587–8**, 610  
and addition by counting 606–7  
dependency goal chunks **601–3**, 607  
hierarchy **596**  
and list activation 597–9  
and the list memory task 593–6  
and production rules 589–90  
slots and values **587**, **594–6**
- Churchland, P.M. 645–6
- Churchland, P.S. 645
- Clark, H.H. 236, 250, 255
- classical view of concepts 158, **169–75**, 187–8  
and borderline cases 172–3  
and definitions **169–70**, 174–5  
and intransitivity of categorization 173–4  
and prototype theories 176–7, 179  
and typicality ratings **170–2**  
and well-defined categories **189**
- Clifton, C. 224, 225
- closure, and perceptual organization **78**, 79
- clustering, and the primal sketch **95**
- CMM (computational model of the mind) **619–32**  
and the brain 620–1  
and connectionist modelling 621–2, 631–2  
and consciousness 621  
and Fodor’s theory of modularity 639  
and neurobiology 648
- coarticulation, in the speech stream **199**
- cocktail party effect **43**
- cognitive appraisals, and emotion **496–501**
- cognitive architectures 458, 460, **584**, 585–614  
Soar 613  
sub-symbolic **609–10**  
*see also* ACT-R architecture; PDP (parallel distributed processing)
- cognitive correlates of consciousness **546**
- cognitive modelling 458, 460, 579–616  
and category-specific impairments 4–5  
defining **579**  
evaluating 611–13  
high and low-level approaches to 579
- hybrid models **585**  
and the Newell Test 611, **612–13**
- rule-based systems **583–4**
- sub-symbolic models 609–10  
*see also* ACT-R architecture; PDP (parallel distributed processing)
- cognitive modules  
and access consciousness **547–8**  
cross-talk between 568–9
- cognitive neuropsychology, and the past-tense debate 626–9
- cognitive psychology  
assumptions made in 7  
defining 2–8  
history of 8–19  
methods 3–6  
and other disciplines 1  
and other sub-disciplines of psychology 1  
sub-disciplines of 3
- cognitive systems, of face recognition **139**
- coherence  
models of memory 537  
and working memory 327–8  
and written language **232–3**, 234
- cohesion, and written language **232**, 233
- cold emotions 482, 483
- Collins, A.M. 215, 279
- colour  
categories 172, 173  
and Marr’s theory of perception 91–2
- Coltheart, M. 49, 54, 208, 209, 629, 641–2, 645
- common ground, in language production **249**, 250–1
- communal self-focus 524
- communication  
and emotion 479  
and group decision making 259–60
- comparability, and SEU (subjective expected utility theory) **387**, 388
- competition, and attention 65
- competitive theories of reasoning **448–9**
- complex actions, and behaviourism 12–13
- complex concepts 166–7  
and prototype theory 180  
and ‘theory’-theory 184
- complex problem solving 365–71  
chess studies 365, 366–7
- component axes, of 3D objects **127–9**
- compositionality

- and connectionist modelling **631–2**
- constraint **649**
- computation** **24–6**
- computational level of cognition **27**
  - and Marr's theory of perception **91, 96**
- computational model of the mind *see* CMM
- (computational model of the mind)
- computer modelling
  - structure mapping engine (SME) **364–5**
  - and thinking **346**
  - and working memory **267**
    - the phonological loop **323, 331–2, 333–4, 335–6**
- computers
  - and emotion **463**
  - and the mind **16–18, 20**
  - and models **20**
- see also* Artificial Intelligence (AI)
- concavity, of 3D objects **127–8**
- concepts **158–9, 163–95**
  - and categories **163–4**
  - and categorization behaviour **164–6**
  - classical view of **158, 169–75, 176–7, 179, 187–8, 189**
  - and cognition **167–8, 191**
  - complexity of **166–7**
  - essentialist theories of **158, 184–7, 188, 190**
  - prototype theory of **158, 175–80, 181, 187–8, 189**
  - 'theory'-theory of **158, 180–4, 185, 188, 190**
- see also* categorization; complex concepts
- conceptual implicit memory **286–8**
- conclusions, and reasoning **418**
- conditional inference **346, 427–37**
  - abstract **427–31**
  - everyday reasoning and the suppression effect **432–7**
- conditionals
  - and deductive reasoning **421**
  - and mental logic **428–9**
  - and mental models **429, 430, 435–6, 440**
  - probabilistic approach to **431**
  - and the probabilistic approach to reasoning **425–6**
- see also* bi-conditionals
- conditions, and production rules **583**
- configural processing **152**
- confirmation bias, and the abstract selection task **438**
- conjunction fallacy **405, 407–8**
- connectionist models **25–6, 29, 618, 628–32**
  - and the CMM **621–2, 631–2**
  - and double dissociation **628–9**
  - language processing **159, 204–6**
  - of the past tense **624–6, 628–9**
  - recognition **216, 460**
  - rules in connectionist networks **629–30**
  - structure and compositionality **631–2**
- see also* PDP (parallel distributed processing)
- connectives
  - and deductive reasoning **421, 422**
  - and the mental models theory of reasoning **425**
- conscious awareness, and attention **66–7**
- consciousness **1, 6, 31, 458, 459–60, 545–77**
  - access consciousness **546, 547–8**
  - altered states of **569–71**
  - automatic processing **545, 560–2, 564**
  - and behavioural control **564–8**
  - and behaviourism **10, 12, 550, 551**
  - biological aspects of **545**
  - cognitive correlates of **546**
  - and cognitive psychology **550–1**
  - cognitive theories of **571–4**
  - and the computational model of the mind **621**
  - controlled processing **545–6, 560–2, 564, 565**
  - and cross-talk between cognitive modules **568–9**
  - defining **546–8**
  - 'easy problems' of **548, 574**
  - and emotion **468, 481**
  - and errorless learning **554, 555, 565, 567–8, 569, 574**
  - explanatory gap **548**
  - as a global workspace **571, 573**
  - 'hard problems' of **548, 574**
  - and implicit learning **459, 555–60, 564, 569**
  - and implicit memory **552–5, 564**
  - and introspectionism **8–9**
  - multiple drafts theory of **572**
  - neurophysiology of **545, 546, 562–3**

- phenomenal consciousness 546, **547–8**  
 philosophical approach to 454, 547,  
**548–50**  
 and PTSD (post-traumatic stress disorder)  
**536**  
 stream of 547, 555, 563  
 and working memory 551, 560, 571–3, 574  
 consensus, and dialogue 255–6  
 conservatism, in judgements **402**, 403, 404  
 consistent mapping condition **561–2**  
 constraint relaxation, in problem solving  
**361–2**  
 constructivist approaches to perception  
**98–101**, 104, 105, 106  
   and perceptual hypotheses **99–101**  
 content words **211**  
 context-sensitivity, typicality effects 179  
 contour generators, and 3D objects **125**, 126,  
 127  
 convexity, of 3D objects **128**  
 Conway, M.A. 297, 298, 507, 512, 517–18,  
 521, 522, 528, 532  
 correspondence models of memory 537  
 Corsi span 315  
 Corteen, R.S. 44  
 Cosmides, L. 444, 447  
 covert recognition **145**, 147  
 Cowan, N. 316, 330  
 Craik, F.I.M. 270–2, 284, 299, 310  
 cross-modal priming **203**  
 culture, and emotions 464–5, 470–2, 474  
 Cummins, D.D. 435  
 curvilinear aggregation, and the primal  
 sketch **95**  
 Cutler, A. 200  
 cybernetics 15
- DA (denying the antecedent) **423–4**  
   and the abstract conditional inference task  
 427, 428, 429, 430, 431  
   and the abstract selection task 438, 439, 440  
   suppression **432**, 433, 434–5, 436
- Damasio, A.R. 480
- Daneman, M. 311–13, 323–4
- Darwin, Charles 463, 469, 479
- Davies, M. 629–30, 645
- Davies, S.P. 359–60
- De Beni, R. 324
- De Groot, A.D. 366, 371
- De Renzi, E. 315
- decision analysis 383, **385–7**
- decision making 6, 344–5, 382–99
- and behavioural decision research 392–6
- and consciousness 547
- descriptive theories of **383**, 396–9
- fast and frugal theories of **410–12**
- influence from outside psychology on 382
- and judgements 382, 400
- mistakes in 382
- normative theories of **383**, 384–96
- prescriptive approach to **383**
- and problem solving 347–8
- prospect theory **396–9**
- and reason-based theories of choice **395**
- see also* judgement
- decision trees 383, 385–7
- declarative memory **282–3**, 290, 299  
   and ACT-R models **586**, 587–8, 590, 591,  
 599, 600, 606–7, 609  
   and past-tense learning 609
- see also* chunks in ACT-R models
- declarative procedural hypothesis, and the  
 words and rules theory 623
- decompositional approach, to morphology  
**214**
- deductive reasoning 421–4, 449, 451
- Deeprose, C. 559
- defeasible inferences **236**, **432**
- definitions, and the classical theory of  
 concepts **169–70**, 174–5
- Dehaene, S. 573
- Dennett, D. 562, 572
- deontic selection task 443–8  
   and deontic conditionals **443**  
   and evolutionary psychology 446–8  
   and indicative conditionals **443**  
   and mental logic 444  
   and mental models 444–5  
   and permission rules **443–4**  
   and the probabilistic approach 445–6
- dependency goal chunks **601–3**, 607
- depolarization, and Marr's theory of  
 perception **92**
- depression  
   clinical depression and memory bias 483,  
 485  
   and consciousness 574  
   and mood congruent memories 490  
   and self-serving biases 493
- depth of processing, in memory **270–2**
- derivational morphology **214**
- Descartes, René 548–9

- descending pathways, and neurobiological theory 645–6
- descriptive clauses/sentences 421
- descriptive theories of decision making 383
- prospect theory 396–9
- Deutsch, D. 44
- Deutsch, J.A. 44
- Devlin, J.T. 4
- Di Lollo, V. 52, 106, 107
- diagnosticity, and Bayes' Theorem 402
- dialogue 231, 245, 253–8
- and consensus 255–6
  - and group decision making 258–60
  - and representational alignment 256–7, 258, 260
  - and routinization 257–8
- Diamond, R. 149–51
- dichotic listening 41, 43, 44, 54, 60, 61
- direct perception 81
- direct retrieval 525, 527
- discourse, and written language 232–44
- discursive approach to categorization 165
- distorted faces, and the inversion effect 151–2
- distraction 59–62
- attending across modalities 61
  - effects of irrelevant speech 60–1
- domain specificity
- and Coltheart's view of modularity 641–2
  - and Fodor's theory of modularity 633, 634, 637
- dominance, and SEU (subjective expected utility theory) 387, 388
- Donaldson, W., detection model of memory 297, 298
- dorsal stream
- and attention 64
  - and face recognition 145–7
  - and perception 102–4, 104–5, 105–6
- dot probe task 489
- double dissociation 627, 628–9, 641
- Draper, S.W. 375
- DRC model of word recognition 208–9, 617, 618, 619–20, 629–30
- Driver, J. 61, 63
- driving, and mobile phones 61
- drugs, and consciousness 569
- dual-task paradigm 310–13
- DuCharme, W.M. 403
- Dunbar, K. 364
- Duncan, J. 56, 57
- Duncker, K. 350, 353, 355–6, 363
- dyslexic children 320
- eavesdropping on the unattended message 43–4
- echoic memory 41
- Eco, Umberto 166
- ecological approach to perception 81–2, 105
- Edelman, S. 134
- Edwards, D. 165
- Edwards, W. 392, 393, 400, 403, 412
- Ehlers, A. 532, 533
- Eich, E. 488, 553
- Einstein, Albert 220
- Einstein, G.O. 276
- Ekman, Paul 469–72, 496
- elaboration, in problem solving 361
- elaborative inferences 236–7
- eliminative materialism 646–8
- Ellis, H.D. 145
- embodied cognition 238–40
- emergent properties, in PDP models 583
- EMG (electromyography) 467
- emotion 1, 6, 31, 458–9, 463–506
- and attention 488–91
  - basic emotions 469–73, 474, 502
  - and behaviourism 463
  - Big Five emotions 469, 470
  - bodily responses to 465–7, 468, 481
  - and consciousness 468, 481
  - dimensional approach to 474–6
  - emotional behaviour and expression 464–5, 468
  - in everyday life 463
  - and face recognition 117–18
  - and facial expression 465, 468, 470, 471
  - feeling emotions 468
  - function of 476–81
  - measuring using psychophysiology 466–7
  - and memory 477, 483–8
  - processing vs manifestation of 482–3
  - and PTSD (post-traumatic stress disorder) 531, 532, 533, 535
  - and semantic interpretation 491–3
  - theories of cognition and 494–502
  - trait and strait 481–2
  - verbal labels for emotions 473–4
- Emotional Stroop task 58, 488–9
- encoding in memory 266, 269, 270–7, 292
- and Fodor's theory of modularity 636

- and implicit learning 560  
 levels of processing 270–3  
 and PTSD (post-traumatic stress disorder) 532, 534  
 relational and item-specific processing 274–7  
 encoding specificity 284  
 Engle, R.W. 313  
 Enns, J.T. 96, 106, 107  
 episodic buffer 328, 572  
 episodic information, and visual attention 49, 54, 56  
 episodic memory 49, 167–8, 269, 290, 299, 617, 618  
     and autobiographical memory 512, 518, 520–2, 526, 527–8, 528–9  
     and PTSD (post-traumatic stress disorder) 532–3, 535, 536  
     and remember judgements 295, 296, 297, 298  
     systems 278, 279–82  
 Erev, I. 409  
 Ericsson, K.A. 369, 370  
 ERPs (event-related potentials) 64–5  
     studies of semantic processing 217, 219  
 errorless learning 554, 555, 565, 567–8, 574  
 essentialist theories of categorization 158, 184–7, 188, 190  
 evaluability principle, in decision making 395–6  
 event-related potentials (ERPs) 64–5  
 everyday behaviour, cognitive processes involved in 6–7  
 Evett, L.J. 47, 48, 49  
 evolutionary psychology 10  
     and the deontic selection task 446–8  
     and emotions 477  
     evaluating 451  
 executive processes, in working memory 323–8, 336  
 experimental methods, and category-specific impairments 4  
 expert problem solving 365  
     and individual differences 372–6  
     role of knowledge in 366–70  
     and the transfer of expertise 371–2  
 explicit memory 553, 555, 559  
 explicit motives, personality and memory 523–4  
 exploratory procedures, and object recognition by touch 119  
 external problem representations 357–8  
 the eye, and perception 74–5, 102  
 eye movements in reading 210–13, 234, 235  
 Eysenck, M.C. 273  
 Eysenck, M.W. 273, 492  
 face recognition 116–18, 134–52, 579, 618–19  
     Bruce and Young model of 139–41, 617, 618, 619  
     and Capgras delusion/syndrome 73, 145–7, 479–80  
     and cognitive architectures 610  
     and Coltheart’s view of modularity 641  
     connectionist model of 141–3  
     and E-FIT images 117  
     and emotion 117–18  
     errors in 138–9  
     familiar and unfamiliar faces 117, 135, 136–8, 152  
     and Fodor’s theory of modularity 636  
     functional model of 139–41  
     and the inversion effect 149–52  
     and neurobiological theories 644–5  
     neuropsychological evidence for 144–8  
     and newborn babies 148–9  
     and object recognition 118, 135–6, 148  
 face recognition units (FRUs) 139, 140, 141, 142, 143, 618  
 facial expression, and emotion 465, 468, 470, 471  
 Falkenhainer, B. 364  
 false frame effect 553–4  
 fan effect, in memory 280, 598  
 fast and frugal heuristics 410–12  
 Fay, N. 259–60  
 featural processing 152  
 featural theory of semantic representation 215–16  
 feature integration theory 56–9  
     and the ‘flanker’ effect 57–9  
     and non-target effects 56–7  
 feature recognition theory, and object recognition 121, 122  
 feedback  
     loops 14–15, 16  
     and object recognition by touch 118  
 Ferreria, F. 224, 225  
 ffytche, D.H. 563, 573  
 Fiddick, L. 448  
 filtering, and the attenuation process 43, 44

- fired production rules **583**  
 first-order relational properties, and face recognition **149**  
 fixations **210–13**  
 ‘flanker’ effect **57–9, 488**  
 flashbulb memories **507–8, 531**  
 fleshed-out mental model  
   and the abstract conditional inference task **430**  
   and the abstract selection task **440**  
   and the deontic selection task **445**  
   and the suppression effect **435**  
 ‘flight or fight’ response **465, 478**  
 flow patterns, in the ambient optic array **87–9**  
 fMRI (functional magnetic resonance imaging) **21, 644**  
   and attention **62**  
   and category-specific impairments **4**  
   and consciousness **563, 573**  
   and emotion **472–3**  
   and face recognition **148**  
   and the phonological loop **322**  
 Fodor, J.A. **22, 645, 646**  
   theory of modularity **460, 632–9, 640–1, 642**  
 formal rule theories of reasoning **425**  
 forward recall, and ACT-R models **592–3, 599, 600**  
 forwards search, in problem solving **358**  
 foveal region of the retina **212**  
 fractionation  
   and the mind **655**  
   and working memory **326–7**  
 frame judgement task **553–4**  
 frame problem, and Artificial Intelligence (AI) **638–9**  
 framing effects **399**  
 free recall, and ACT-R models **592, 600**  
 French, C.C. **492**  
 Frensch, P.A. **375**  
 frequency of events, and the conjunction fallacy **407–8**  
 Freud, S. **550**  
 FRUs *see* face recognition units (FRUs)  
 full primal sketch, and Marr’s theory of perception **94, 95, 96**  
 full-listing approach, to morphology **214**  
 function words **211**  
 functional accounts of cognition **29**  
 functional amnesia **534**  
 functional fixity **355–6**  
 functionalism, and consciousness **549–50**  
 functionalist psychology **10**  
 functions, fractionation of **34**  
 fuzzy categories, and prototype theories **190**  
 Gabrieli, J.D.E. **290**  
 Gaeth, G.J. **242**  
 galvanic skin response (GSR) **44, 144**  
 gambles, and decision making **384, 392–3, 398, 399**  
 gambling task **480**  
 garden-path sentences **223, 224–5, 309–10**  
 Gardiner, J.M. **297**  
 Garrod, S. **237, 254**  
 Gathercole, S.E. **58, 59, 330**  
 Gaussian blurring, and Marr’s theory of perception **93–4**  
 Gelman, S. **185, 187**  
 general events, and autobiographical memory **517–18, 526, 528**  
 generalization, and ACT-R models **597**  
 generalized cones, and object recognition **124, 125, 132, 134**  
 generative retrieval **525–7, 528, 529**  
 generativity, and the working self **523**  
 genetic engineering **22**  
 Gentner, D. **364**  
 Gentner, D.R. **364**  
 geons, and object recognition **132, 134**  
 Gerhardstein, P.C. **134**  
 Gestalt psychology **9–10, 13**  
   and simple problem solving **353–6, 361**  
 Gestalt theory of perception **78–80**  
   and affordances **89–90**  
   and the primal sketch **95**  
 Gibson, J.J.  
   theory of perception **80–90, 101, 105, 106, 109, 617**  
     and action **80–1, 118**  
     affordances and resonance **89–90, 104–5, 113**  
     and the ambient optic array **82–9**  
     ecological approach **81–2, 105**  
     and Marr’s theory **90–1**  
     and visual illusions **98**  
 Giesbrecht, B. **52**  
 Gigerenzer, G. **406, 407, 409, 410–11, 412**  
 Gilhooly, K.J. **352, 359, 373**  
 Glenberg, A.M. **238, 239–40, 272–3, 276**  
 Glick, M.L. **363**

- global workspace, consciousness as a 571, 573  
 Glucksberg, S. 172–3, 180–1, 235  
 goal attainment, memories of 518  
 goal hierarchy, and the working self 510–11, 525  
 goal specificity, and problem solving 374–5  
 goal stack 586, 587, 610  
 goal transformations, and ACT-R models 596–7  
 goals  
   and ACT-R models 590–1  
   emotion and readjusting 477, 478  
   and the working self 522–5  
 Gold, I. 645, 646, 647  
 Goldberg, N. 590, 591  
 Goldman, L. 165  
 Goldstein, D.G. 410–11, 412  
 Goldstone, R.L. 183  
 good continuation, and perceptual organization 78  
 Goodale, M.A. 103, 104  
 Goodman, Nelson 181  
 graceful degradation, in PDP models 582–3  
 Graf, P. 290–1  
 Grainger, J. 207  
 grammar  
   artificial grammar and implicit learning 557–9  
   and speech errors 247, 248  
   and syntax 220–2  
 Gray, J.A. 66  
 Green, A.J.K. 373, 374, 375  
 Greer, M.J. 4–5  
 Gregory, Richard 99, 100, 101  
 Grether, D.M. 393–4  
 grey level description, and Marr's theory of perception 91–2, 618  
 Grey, N. 537  
 Grossman, M. 281  
 group decision making, dialogue and monologue distinction in 258–60  
 grouping, and the primal sketch 94, 96–7  
 groups, and chunks in ACT-R models 593–4  
 GSR (galvanic skin response) 44, 144, 467, 500  
 Habermas, T. 519  
 Hackman, A. 536  
 Haider, H. 375  
 Halligan, P.W. 63  
 HAM (human associative memory) 585, 587  
 Hampton, J.A. 173–4, 180–1  
 hand clap, waveform 38  
 haptic information 119–20  
 haptic perception (touch) 74, 77  
 Haque, S. 528  
 Hardcastle, V.G. 574  
 Harris, M. 370  
 Hartsuiker, R.J. 252  
 Hasher, L. 286  
 Haviland, S.E. 236  
 Hay, D.C. 139, 141  
 Hayes, J.R. 25, 356–7  
 hazard management, and evolutionary psychology 451  
 Hermann, D.J. 368  
 heuristics *see* mental heuristics and biases  
 hierarchy, in the list memory task 596  
 Hildreth, E. 93, 96  
 history of cognitive psychology 8–19  
 Hitch, G.J. 310–11, 312, 316, 323, 325, 333–4, 335, 594  
 Hixon symposium 12, 14  
 hobbits and orcs task 357, 359, 360  
 Hoffrage, U. 408  
 Holding, D.H. 371  
 Holmes, A. 536  
 Holyoak, K.J. 363  
 homographs, interpretation and emotion 492  
 homophones 210  
   and implicit memory 553  
   interpretation and emotion 491–2  
 homunculus problem, and consciousness 572  
 Hor, mask of 100  
 horizon ratio relations, and the ambient optic array 82–4  
 Horton, W.S. 251  
 hot emotions 482, 483  
 Hsee, C.K. 395–6  
 Humphreys, G.W. 47, 48, 49, 64  
   and object recognition 114–15, 131, 135  
 Humphreys, J.W. 56, 57  
 Hunt, R.R. 273–4, 276, 289  
 Hupe, J.M. 107  
 hybrid models 585  
 hypnosis 66, 569–70  
 hypotheses, perceptual 99–101, 107

- IAC (interactive activation and competition network) models  
 face recognition 141–3, 144, 147, 204  
 language processing 159  
 visual word recognition 206–7, 215
- IAM (incremental analogy machine) 365  
 iconic memory 47  
 impasse, in problem solving 361  
 implementational level of cognition 29  
 implicit learning 459, 555–60, 569  
 and speech segmentation 202  
 implicit memory 286–92, 459  
 and amnesia 290–1, 554, 555  
 and consciousness 552–5  
 memory systems accounts of 290  
 perceptual and conceptual 286–8  
 TAP account of 288–9, 291
- implicit motives, personality and memory 523–4  
 implicit negation 439  
 impoverished figures, and perceptual hypotheses 99–100  
 incremental, parsing as 222  
 incremental analogy machine (IAM) 365  
 incremental interpretation 235  
 independence, and SEU (subjective expected utility theory) 387, 388, 390–1  
 indicative conditionals 443  
 indirect perception 81  
 individual differences  
   and phonological working memory 317  
   in problem solving 344, 372–6  
   in reasoning 451–2  
   and vocabulary acquisition 330  
   and working memory 267  
 inertia effect, and Bayes' Theorem 402–3  
 inference-making 236–7  
 inferences, logically valid 423–4  
 inflectional change 214  
 inflectional morphology 214  
 information, emotions as 479–80  
 informational encapsulation  
   arguments against 639–41, 646  
   and Fodor's theory of modularity 633, 634–6  
 information ‘flow’, and perception 75–6, 101  
 information gain 441, 442  
 information processing approach, to problem solving 358–62  
 information transfer, and dialogue 256, 260  
 initial mental model, and the abstract conditional inference task 429  
 inner-loop monitoring of speech 252–3  
 insight, in problem solving 353–4, 361–2, 376  
 integration masking 50  
 integrative theories of reasoning 448, 449, 452–3  
 intentional nature of mind 655  
   aboutness 23  
 interactive alignment, and dialogue 256  
 interactive activation and competition network models *see* IAC (interactive activation and competition network) models  
 interactive view, of semantic ambiguity 217–18  
 internal problem representations 357–8  
 internal structure, of categories 171–2  
 interruption masking 50  
 intransitivity of categorization 173–4  
 introspectionism 8–9, 11, 550, 551  
 invariance, and SEU (subjective expected utility theory) 387, 388  
 invariants, and the ambient optic array 82–6  
 inversion effect, and face recognition 149–52  
 IQ, and logical reasoning 346, 451–2  
 irrelevant speech effect 60–1, 320, 328  
 Isaacs, E.A. 250  
 isomorphic problems 357  
 isotropic scientific confirmation 637–8, 642, 646  
 item-specific processing, in memory 274–7
- Jacobs, A.M. 207  
 Jacoby, L.L. 286, 553  
   process-dissociation framework 292–4, 299  
 James, William 10, 308, 499  
   and consciousness 547, 550, 555, 571  
   and the James-Lange theory of emotion 494, 495, 496, 497  
 Johnson, M.K. 280, 286  
 Johnson-Laird, P.N. 477  
 Jones, D.M. 60, 327–8  
 Joordens, S. 54  
 judgement 344–5, 400–10  
   and Bayes' Theorem 345, 346, 400–2, 404  
   and choice 382, 400  
   and decision making 382, 400  
 errors of 383

- heuristics and biases 404–8
- overconfidence in 408–10
- and problem solving 347–8
- see also* decision making
- Juslin, P. 409, 410
- Just, M.A. 310
- Kahneman, D. 397, 399, 402, 404–5, 406, 407–8, 413
- Kanizsa's illusory square 71, 72, 98
- Kaschak, M.P. 239–40
- Keil, F. 183
- Kellenbach, M.L. 217
- Kelter, S. 257
- Kemp, R. 137
- Keren, G.B. 409
- Keysar, B. 251, 252
- Kilgour, A.R. 137
- kinesthesia, and object recognition by touch **118, 119**
- Klatzky, R.L. 119
- Knoblich, G. 362
- knowing, remember and know judgements 295–9
- knowledge
  - autobiographical 517–20, 521
  - and categorization 158
  - meaning and embodiment 237–40
  - prior knowledge and reasoning 420
    - relating language to 237
- Knowlton, B.J. 558
- Koffka, K. 79
- Korsakoff's Syndrome 281
- Kripke, S.A. 174, 175, 185
- Kroska, A. 183
- Külpe's Würzburg school of introspection 9
- Kunst-Wilson, W.R. 556
- LaBerge, D. 57
- Lakoff, G. 235–6
- Lang, Peter 474
- Langdon, R. 645
- Lange, Carl, and the James-Lange theory of emotion 494, 495, 496, 497
- language 1, 6, 19, 159–60
  - in action 160, 231–64
  - and the algorithmic level of cognition 29
  - and behaviourism 12–13
  - and categorization 158, 165, 189
  - and communication 231
  - comprehension 159–60, 579
- and connectionism 25, 26
- and dialogue 160
- and problem solving 347
- production 160, 245–53
  - and dialogue 253–8
  - message selection and audience design 249–51
  - and speech errors 245–9
- and symbol systems 25–6
- understanding 7, 160
- see also* speech; written language
- language processing 159, 197–230
  - IAC model of 159
  - and knowledge
    - bottom-up (autonomous) view of 159
    - top-down (interactive) view of 159
  - and the mental lexicon 159, 160, 197, 198, 213–19
  - see also* sentence comprehension and processing; word recognition
- language of thought hypothesis 649
- Lashley, Karl 12–13
- lateral geniculate nucleus (LGN) 102
- Law of Pragnanz **79**
- Lawson, R. 131
- Lazarus, Richard, cognitive appraisal theory of emotion 499, 500–1, 502
- learning
  - and ACT-R models 601–8
  - and consciousness 550
  - errorless 554, 555, 565, 567–8, 569, 574
    - implicit 202, **555–60**, 569
- Lebiere, C. 585, 591, 610, 612–13
- Lederman, S.J. 119, 137
- LeDoux, Joseph 494, 501–2
- Leipzig school of introspection 8–9
- lesioned animals and emotion **501–2**
- level-dependent explanations of cognition 27–30
  - algorithmic level 27–9
  - computational level 27
  - implementational level 29
- Levelt, W.J.M. 257
- Levin, I.P. 242
- Levine, J. 548
- lexical competition, and spoken word recognition 204–6
- lexical concepts 168
- lexical decision tasks **43**, 626–8
  - and information encapsulation 635–6

- lexical effects, in visual word recognition **207**  
 lexical models of segmentation **199**  
 LGN (lateral geniculate nucleus) 102  
 Liberman, K. 315  
 Lichtenstein, S. 393  
 Lieberman, M.D. 555  
 life story **519–20**, 522  
 lifespan retrieval curve 509–10, **511–14**  
   childhood amnesia 511–12, 513  
   recency period 511, 513–14  
   the reminiscence bump 511, **512–13**  
 lifetime periods **518–20**, 522, 526, 528  
 likelihood ratio, and Bayes' Theorem **401**  
 linguistic impairments, and the past-tense debate 626–9  
 lip-reading 61  
 list memory **592–9**, 600  
 Lockhart, R.S. 270–2, 284, 299, 310  
 Loftus, E.F. 215  
 logic  
   and reasoning **345**, **418–19**  
     deductive reasoning 421–4  
     form and meaning in logic 424  
   logically valid arguments **421–3**  
   logically valid inferences 423–4  
 Logie, R.H. 316  
 long-term memory (LTM) 266, 269–306, 307  
   and the central executive 323  
   and consciousness 560  
   encoding 266, **269**, 270–7, 292, 299  
   implicit 286–92  
   and Jacoby's process-dissociation framework 292–4, 299  
   multiple memory systems 278–82  
   and protocol analysis 350  
   and reasoning 436  
   remember and know judgements 295–9  
   retrieval 266, **269**, 284–6, 292–4, 299  
   and short-term memory 308–10, 330  
   storage **269**, 277–83  
   *see also* episodic memory; semantic memory  
 Lopes, L.L. 412  
 loss aversion **398**  
 low-frequency words **208–9**  
 LTM *see* long-term memory (LTM)  
 Lucas, M. 217  
 Luchins, A.S. 354  
 Luchins, E.H. 354  
 Luzzatti, C. 63  
 Lynch, E.B. 190  
 McAdams, D.P. 523  
 McCarthy, John 17  
 McClelland, J.L. 206, 209, 282, 609, 622, 624, 625  
 McCloskey, M. 172–3, 180–1  
 McDaniel, M.A. 273–4  
 McDermott, K.B. 286, 289  
 MacDonald, M.C. 223  
 Macintyre, A. 167  
 McKee, R. 554  
 Macken, W.J. 328  
 MacLeod, C. 489, 490  
 McQueen, J.M. 206  
 Magritte, René 81  
 Malt, B.C. 185, 186, 189  
 Mandler, G. 274, 275–7, 292  
 Marcel, A.J. 551, 554, 556  
 Marchman, V. 609  
 Markus, H. 523  
 Marr, David 24  
   levels of cognition 27–30, 346, 461  
     and categorization 159  
     and cognitive modelling 583, 585, 609  
     cognitive psychology and the brain 643–4  
   perception theory 90–8, 101, 104, 109  
     and the 2½D sketch 91, 95–6, 97, 115, 123, 124  
     and 3D object-centred description 91, 97, 123, 124  
     evaluating 96–7  
     and grey level description 91–2, 618  
     and object recognition 91, 97, 104, 114–15, 123, 124–31, 132, 134  
     and the primal sketch 91, 92–5, 96–7, 114  
     and visual illusions 98  
 Marslen-Wilson, W.D. 202–3, 204, 214, 222, 626–7, 628, 639–40, 641  
 masking 48–9  
   and attention 54  
   backward **47**, 49, 50, 106–8  
 matching bias **438**  
 matching effect, in Wason's selection task **438**, 439, 440, 442  
 Mathews, A. 490  
 MCM (mood congruent memory) **483–4**, 485, 488, 490

- MDM (mood dependent memory) **484–8**  
means-ends approach, to problem solving **358–9**, 368, 374  
medical diagnosis, and protocol analysis 352  
Medin, D.L. 181–2, 184, 190  
memory 1, 6, 265–341  
and the algorithmic level of cognition 29  
and chess skill 371–2  
and consciousness 550  
correspondence and coherence models of 537  
echoic **41**  
and emotion 477, 483–8  
and Gibson's theory of perception 90, 105  
iconic **47**  
mental representations of 617  
as a multifaceted system 307–8  
and problem solving 347  
and symbol systems 25  
*see also* autobiographical memory; episodic memory; implicit memory; long-term memory; procedural memory; semantic memory; short-term memory (STM); working memory  
Mendel, Gregor 11  
mental heuristics and biases **404–8**  
evaluating 406–8  
fast and frugal heuristics **410–12**  
mental illness, diagnosis of 167  
mental lexicon 159, 160, 197, 213–19, 617  
accessing word meanings 215–19  
and information encapsulation 636  
and morphology 213–15  
and word recognition 198, 207, 640  
mental logic theories of reasoning **424**, 425, 451  
and the abstract conditional inference task 428–9, 439  
and the abstract selection task 438–9  
and the deontic selection task 444  
evaluating 449, 450  
and the suppression effect 434–5  
mental models theory of reasoning **424**, 425, 451, 452  
and the abstract conditional inference task 429–30  
and the abstract selection task 439–40  
and the deontic selection task 444–5  
evaluating 449–50  
and the suppression effect 435–6  
mere exposure effect **499–500**  
Merikle, P.M. 47–8, 54  
Mervis, C.B. 171  
metacognitive processes, and problem solving **373**  
metaphors, and text interpretation 235–6  
Metcalfe, J. 488  
metrical foot **199–200**  
Meyer, D.E. 554  
Miller, G.A. 15–16  
Milner, A.D. 103, 104  
Milner, B. 281  
mind/body dualism 548–9  
Miyake, A. 327  
mobile phones, and driving 61  
modal emotions 474  
modal (multi-store) memory model **270–2**, 309, 310  
modalities, attending across 61  
models 19–20  
Modolo, K. 242  
modular input systems, and Fodor's theory of modularity 633–4, 636  
modularity 632–43  
debates about 639–42  
Fodor's theory of 632–9, 640–1, 642  
operational definition of 632–3  
modularity of cognition 161, 655  
modules, and Marr's theory of perception 91–2, 95, 97  
monism 549–50  
monologue 245  
and group decision making 258–60  
morphemes **213**, 214–15, 220  
morphology 213–15, 220  
Morris, C.D. 284, 285, 288  
Moses illusion sentences 241, 242  
motion  
and Gibson's theory of perception 86–9  
perception of 9  
motion parallax **87**  
Motley, M.T. 252  
MP (modus ponens) **422–3**, 424  
and the abstract conditional inference task 427, 428, 429, 430, 431  
and the abstract selection task 438, 439, 440  
suppression **432**, 433, 434  
MRI studies *see* fMRI (functional magnetic resonance imaging)  
MT (modus tollens) **423**, 424  
and the abstract conditional inference task 427, 428, 429, 430, 431

- and the abstract selection task 438, 439, 440  
 suppression 432, 433, 434  
 and Wason's selection task 439
- Müller-Lyer illusion 71, 72, 98–9, 383
- multi-store (modal) memory model 270–2
- multiple drafts theory of consciousness 572
- Murphy, A.M. 403
- Murphy, G.L. 165, 181–2, 183, 184
- Murphy, S.T. 565–6
- music, and distraction 59, 61
- Myung, I.J. 611
- naming, and recognition 115–16
- necessary inferences 236
- Necker cube 71, 72, 98
- neighbouring words, visual recognition of 209
- Neumann, R. 555
- neural network modelling *see* PDP (parallel distributed processing)
- neurobiological theories
- and cognitive theories 461, 644–6
  - and the radical neuron doctrine 646–8
- neuroimaging methods *see* brain-imaging techniques
- neuropsychology *see* cognitive neuropsychology
- Newell, A. 25, 584
- Newell Test 611, 612–13
- Newton, Sir Isaac 11, 14, 19
- Nichelli, P. 315
- 9-dot problem 355, 356
- Nishihara, H.K., Nishihara and Marr's theory of object recognition 124–31
- Nissen, M.J. 558
- nodes, in PDP models 580–1
- non-adversary problem solving 365
- non-literal meaning, and text interpretation 235–6
- non-modular central systems, and Fodor's theory of modularity 633, 636, 637–9
- non-target effects, and feature integration theory 56–7
- nonaccidental properties, in object recognition 132–4
- normal cognition, and category-specific impairments 3–5
- Norman, D.A. 44, 325, 336, 357, 572
- Norman, J. 104, 105
- normative theories of decision making 383, 384–92, 411
- and decision analysis 383, 385–7
- normative theory of subjective expected utility 345
- Norris, D. 220
- Norris, D.G. 334, 335
- novelty hypothesis 513
- NRUs (name recognition units) 141, 142
- Oaksford, M. 445
- Oatley, K. 477
- object recognition 115–16, 118–35
- by touch 118–20
  - and Coltheart's view of modularity 641
  - and face recognition 118, 135–6, 148
  - and Fodor's theory of modularity 636, 639
  - and Marr's theory of perception 91, 97, 104, 114–15
  - object-centred vs viewer-centred descriptions 122–3, 124
  - three-dimensional objects 119, 122–3, 124–35
  - two-dimensional objects 120–2
- object-centred descriptions, and Marr's theory of perception 95, 123, 124
- observer motion, and Gibson's theory of perception 86–9
- occluding contours of an object 125
- occlusion 87
- Ockham's Razor 612
- Ohlsson, S. 361–2
- Ohm's law 620, 629
- operationism 11, 12
- operators, and problem solving 361
- optimal viewing position (OVP) 211–12
- organic amnesia 534
- orthography of a word (spelling) 207–10
- Ortony, A. 184, 474
- Osler, S. 284
- overconfidence in judgements 408–10
- OVP (optimal viewing position) 211–12
- Page, M.P.A. 334, 335
- Pandemonium systems 121
- Papagno, C. 330–1
- parafoveal region of the retina 212
- parallel activation, and spoken word recognition 202–4
- parallel information capture 42
- parallel processing
- and attention 60–1, 65
  - visual information 49, 52, 59
- see also* PDP (parallel distributed processing)

- parallel search, in visual displays 55–6
- parasympathetic ANS (autonomic nervous system) 465, 466
- parsing 220, 222–6
- constraints on 225–6
  - garden path model of 223, 224–5
- partial matching, and ACT-R models 599
- partial report superiority effect 47
- partonomic knowledge structures 519
- past participle 225
- past-tense debate 622–9
- and children's learning 607–8, 608–9
  - and cognitive neuropsychology 626–9
  - and connectionist modelling 624–6
  - and the words and rules model 622–4
- pattern discrimination, and the visual system 103
- pattern recognition, and expert problem solving 369, 370
- Patterson, K. 622, 625
- Paulesu, E. 321, 322
- PDP (parallel distributed processing) 460, 579–83
- and ACT-R architecture 608–10
  - backward propagation of error 582
  - emergent properties 583
  - evaluating 612, 613
  - hidden layer 581, 582
  - input layer 581, 582
  - and list activation 597
  - output layer 581, 582
  - and rule-based systems 583–4
  - threshold value 580, 581
  - training process 581–2
  - units/nodes 580–1
  - weight of links 580, 581
- Pecher, D. 48
- Penrose triangle 100–1
- perception 31, 34, 35, 71–112
- approaches to 75–7
  - and consciousness 459
  - constructivist approach to 98–101, 104, 105, 106
  - dual-process approach to 105–6
  - errors and visual illusions 71–3
  - Gestalt approach to 78–80
  - Gibson's theory of 80–90, 90–1, 105, 106, 109
  - and the human visual system 74–5, 102–8
  - and judgement 383
- Marr's theory of 90–8, 101, 104, 109
- and neurobiological theories 645–6
- perceiving and sensing 73–4
- and problem solving 347
- and reasoning 420
- and recognition 113
- and stored knowledge 35
- and symbol systems 25
- and three-dimensional (3D) objects 73, 80
- and top-down processing 77, 98, 99, 105, 106–8
- and two-dimensional (2D) objects 73
- see also* bottom-up processing
- perceptual organization 78–80
- perceptual classification
- and categorization 158, 167
  - and recognition 114, 115, 135
- perceptual hypotheses 99–101, 107
- perceptual implicit memory 286–8
- perceptual processes 1, 6, 34–5
- see also* attention; perception; recognition
- peripheral (integration) masking 50
- permission rules 443–4
- person identity nodes (PINs), and face recognition 139, 140, 141, 142–3, 147, 618
- personality, and the working self 523–4
- perspective in communication 240, 242–4
- PET (Positron Emission Tomography) 4, 322, 472, 644
- Peterson, C.R. 403
- phrase structure 221
- phrase structure grammar 221
- phenomenal consciousness 546, 547–8
- Phillips, L.D. 387, 403, 409
- philosophy, and consciousness 454, 547, 548–50
- phlogiston 549
- phonemes 199
- errors involving 247, 249
- phonetic feature 204
- phonological loop 319–23, 329–35, 335–6
- development and cross-linguistic differences 319–20
  - and the irrelevant speech effect 320
  - neural basis of the 321–2
  - primacy model 334, 335
  - and serial order in behaviour 332–5
  - two-component model 331–2, 335
  - and vocabulary acquisition 329–31
  - and word-length effects 322–3

- phonological representation **201**  
 phonological working memory **317–23**  
   and articulatory suppression **317–19**  
 phonology of a word (sound) **207–10**  
   and speech errors **247, 248**  
 physics  
   and perceptual processes **34**  
   and representation **23**  
   and the unobservable **10–11, 20**  
 physiological resources, emotions and the mobilization of **478**  
 physiology, and perceptual processes **34**  
 Piaget, J. **418**  
 Picasso, Pablo, *Rites of Spring* **125, 126**  
 Pickering, M.J. **234–5**  
 Pillemer, D. **508, 524**  
 Pinker, S. **175, 622, 625**  
 PINs *see* person identity nodes (PINs)  
 Pitt, M.A. **611**  
 place tokens, and the primal sketch **94, 95**  
 Plato **463**  
 Pleydell-Pearce, C.W. **512, 517–18, 522, 532**  
 Plott, C.R. **393–4**  
 Plunkett, K. **609**  
 poker chip paradigm **400–2**  
 pole, and Gibson's theory of perception **87, 89**  
 Polson, P.G. **370**  
 Poncet, M. **147**  
 Ponzo illusion **634, 635**  
 positional confusions **599**  
 positional hypothesis, and the phonological loop **333–4**  
 post-traumatic stress disorder *see* PTSD (post-traumatic stress disorder)  
 posterior odds, and Bayes' Theorem **401**  
 Potter, J. **165**  
 power law of practice **369**  
 practice  
   and expert problem solving **369–70**  
   and production rules and ACT-R models **590**  
 pragmatic modulation, principle of **435–6**  
 pre-lexical models of segmentation **199, 201**  
 preference reversal phenomenon **393–4**  
 premises, and reasoning **418**  
 prescriptive approach to decision making **383**  
 primacy debate, and emotion **499–500**  
 primacy effect, and ACT-R models **593, 598, 599**  
 primacy model, of the phonological loop **334, 335**  
 primal sketch, and Marr's theory of perception **91, 92–5, 96–7, 114**  
 primary task, and phonological working memory **317**  
 prime words **203**  
 primed lexical decision tasks **626–7, 635–6**  
 priming  
   and the ACT-R model **588**  
   affective **565–6**  
   and auditory attention **43**  
   and consciousness **574**  
   and implicit learning **559–60**  
   and implicit memory **554–5**  
   and unseen information **48–9**  
   and visual displays **57, 58**  
 primitives  
   and 3D objects **129**  
   and Marr's theory of perception **94, 95**  
 Prince, A. **582**  
 principle of pragmatic modulation **435–6**  
 prior odds, and Bayes' Theorem **401**  
 probabilistic approach to reasoning **425–6, 452**  
   and the abstract conditional inference task **431**  
   and the abstract selection task **440–2**  
   and the deontic selection task **445–6**  
   evaluating **450**  
   and the suppression effect **436**  
 problem solving **6, 344, 347–81**  
   and the ACT-R model **604–8, 610**  
   analogical **363–5, 376**  
   ‘complex’ **365–71**  
   defining attributes of problems **349–50**  
   errors in **344, 347**  
   everyday **347, 348**  
   and Fodor's theory of modularity **636**  
   individual differences in **344**  
   and protocol analysis **350–3, 368**  
   and reasoning, judgement and decision making **347–8**  
   research prospects **371–6**  
   ‘simple’ (puzzles) **353–62**  
     Gestalt legacy **353–6, 361**  
     information processing approach to **358–62**  
     representation in **347, 356–8**  
 and state-space **344, 346, 357, 359, 360**  
 and symbol systems **25**

- problem-reduction approach, to problem solving **358–9**
- procedural memory **282**, 283, 290, 299
- in ACT-R models 60–9, 586, 588–90, 603
- process accounts of cognition **29**
- production compilation **586–7**
- in ACT-R models **589–90**, 601–3
  - and past-tense learning 609
- production rules
- and ACT-R models **588–9**, 596–7, 599, 600, 602–3
  - addition by counting 604–7
  - and PDP models **583**
- productivity, and morphology **220**
- progressive deepening **359**
- prominence effect **394–5**
- pronunciation, and word recognition 198, 208–9
- property-listing, and concepts **166**, 171
- proprioception, and object recognition by touch **118**
- prosopagnosia 73, **144–8**, 618–19, 644
- prospect theory **396–9**
- and framing effects **399**
  - and loss aversion **398**
- protocol analysis, in problem-solving research 350–3, 368
- prototype theory of concepts 158, **175–80**, 181, 187–8, 189
- and complex concepts 180
  - and fuzzy categories **190**
  - and typicality effects 177–9
- proximity, and perceptual organization **78–9**
- pseudohomophones **208**
- psychiatric illness, and autobiographical memory 509
- psychogenic amnesia **534**
- PTSD (post-traumatic stress disorder) 459, 489, 508, 510, 522, 529–37
- avoidance symptoms **533–4**
  - hotspot images 537
  - hyperarousal symptoms 534
  - impact of symptoms 534–5
  - memory intrusions 527, 530, 532–3, 535–7
  - reports of the traumatic event 531
  - response at the time of trauma 531–2
  - subsequent psychological symptoms 532–4
- Putnam, H. 174–5, 185, 186
- qualia 547
- Quillian, M.R. 279
- Quinean scientific confirmation **637–8**, 642
- RAA (reductio ad absurdum) **428**, 431
- Rabinowitz, M. 590, 591
- radical neuron doctrine 646–8
- random generation task 326
- Rapid Serial Visual Presentation (RSVP) **50–4**
- rarity assumption, and Wason's selection task **441**
- rational analysis theory **585**
- rationality 346
- and errors 346
- Raymond, J.E. 52
- RB (reminiscence bump) 511, **512–13**
- re-encoding, in problem solving **361**
- re-entrant pathways, and visual masking **107**
- re-entrant processing, and visual masking **107–8**
- reading
- eye movements in 210–13, 234
  - and Fodor's theory of modularity 636, 639
  - reading span task 312–13, 323–5
- reality monitoring **570**
- reason-based theories of choice **395**
- reasoning 6, 345–6, 418–55
- competitive theories of **448–9**
  - and concepts 168
  - conclusions and premises **418**, 422
  - and conditional inference **427–37**
  - deductive 421–4, 449
  - errors in 419
  - in everyday life 419–20
  - and Fodor's theory of modularity 636
  - individual differences in 451–2
  - integrative theories of **448**, 449, 451–2
  - and logic 345, **418–19**, 421–4
  - logical reasoning and IQ 346, 451–2
  - and problem solving 347–8
  - theoretical evaluation 449–51
  - and Wason's selection task 346, 437–48
- see also* mental logic theories of reasoning; mental models theory of reasoning; probabilistic approach to reasoning
- Reber, A.S. 556–8
- recall-reconstruction paradigm, in chess studies 366, 367, 371
- recency effect, and ACT-R models **593**, 598, 599

- recency period 511, 513–14  
 recognition 34, 35, 113–56  
     different types of 115–23  
     and encoding in memory 275–7  
     and language perception 219–20  
     perception for 75  
     process of 113–14  
     and stored knowledge 35  
     *see also* face recognition; object  
     recognition; word recognition  
 recognition heuristic **412**  
 recollective experience 516, **522**  
 reduced relative clauses **224**  
 reductio ad absurdum (RAA) **428**, 431  
 reductionism, neurobiological 647–8  
 referential communication task 250  
 refractory period, and Rapid Serial Visual  
 Presentation (RSVP) **52**  
 regression effect, and overconfidence in  
 judgements 410  
 relational processing, in memory **274–7**  
 remember and know judgements 295–9  
 reminiscence bump (RB) 511, **512–13**  
 Renkl, A. 373  
 Rensick, R.A. 96  
 representation **22–4**  
     in expert problem solving 368–9  
     in puzzle problem solving 347, 356–8  
 representational alignment, in dialogue  
 processing 256–7, 258, 260  
 representativeness heuristic **404–5**  
     and base-rate neglect 406–7  
 requisite decision modelling **387**  
 resonates, and Gibson's theory of perception  
**89–90**  
 the retina  
     and eye movements in reading 212  
     and perception 74–5, 76, 102  
         bottom-up 80, 91  
         Marr's theory of 92, 93, 95  
 retrieval mode **522**  
 Reynolds, R.I. 371  
 Richards, A. 492  
 Richardson-Klavehn, A. 286  
 Riddoch, M.J. 131  
 Rips, L.J. 171, 182, 188, 428, 449  
 risky choice *see* SEU (subjective expected  
 utility theory)  
 Robertson, D.A. 238  
 Rock, Irvin 99  
 Roediger, H.L. 286, 288–9  
 Rosch, E.H. 170–2, 177, 179  
 Ross, B.H. 165, 279, 280  
 Roth, E.M. 179  
 routinization, in dialogue processing **257–8**  
 RSVP (Rapid Serial Visual Presentation)  
**50–4**  
 Rubin, David 512, 513  
 rule-based systems, and PDP models **583–4**  
 Rumelhart, D.E. 206, 609, 624  
 saccades **210**  
 Saffran, J.L. 202  
 St John, M.F. 559  
 Salamé, P. 320, 328  
 Salovey, P. 524–5  
 Salvucci, D.D. 587  
 Samuelson, Paul 398  
 Sanford, A.J. 237, 243–4  
 Sappington, B.F. 165  
 SAS (supervisory attentional system)  
 325–6, 572  
 satisficing methods of problem solving **411**  
 Savage, L. 384, 387, 390  
 SC (skin conductance) 467, 480  
 Schachter, Stanley, and the Schachter-  
 Singer theory of emotion 495, 496–7  
 Schacter, D.L. 278, 281, 286, 290  
 Schaeken, W. 427, 429  
 Schelling, Thomas 395  
 schema, life story **519**, 520, 522  
 Scherer, K.R. 473–4  
     appraisal theory of emotions 497–8  
 Schneider, G.E. 103  
 Schneider, W. 545, 560–2  
 Schober, M.F. 255  
 Schoenfeld, A.H. 368  
 Schraagen, J.M. 372  
 Schroyens, W. 427, 429  
 Schunn, C.D. 372  
 Schvaneveldt, R.W. 554  
 science  
     models and the mind 19–22  
     and representation 22–4  
     and the unobservable 10–11  
     *see also* chemistry; physics  
 scientific confirmation, and Fodor's theory  
 of modularity 637–8  
 SCR (skin conductance response) **144–5**  
 search process, in problem solving **358–61**  
 Searcy, J.H. 151  
 second-order relational properties, and face  
 recognition 149–52

- secondary task, and phonological working memory **317**  
 Sedgwick, H.A. **82**  
 segmenting the speech stream **198–202**
  - lexical models **199**
  - pre-lexical models **199**, **201**
 Seidenberg, M.S. **209**  
 selective processing, in written language **240–2**  
 the self **458**
  - and autobiographical memory **459**
  - self hypothesis **513**
  - self-defining experiences **513**
  - self-defining memories **525**
  - self-defining moments **524**
  - self-explaining effect, in problem solving **373–4**
  - self-monitoring speech **251–3**
  - self-serving attribution bias **493**
  - semantic ambiguity **217–18**
  - semantic classification **167**
    - and recognition **114**, **115**
  - semantic content **213**
  - semantic features **215–16**
  - semantic information, and visual attention **49**, **54**, **56**
  - semantic information units (SIUs), and face recognition **141**, **142**, **143**, **147**
  - semantic interpretation, and emotion **491–3**
  - semantic memory **271**, **278–82**, **290**
    - and concepts **167–8**
    - and episodic memory **520–2**
    - and know judgements **295**, **297–8**
  - semantic network theory of emotion **486–7**
  - semantic organization **213**
  - semantic priming **216–17**
  - semantic representations **215–17**
    - ERP studies of semantic processing **217**, **219**
    - sensation, and perception **73–4**
    - sensory neglect **62–4**, **73**
    - sensory transducers, and Fodor's theory of modularity **633**
    - sentence comprehension and processing **219–26**
      - cleft sentences **242**
      - descriptive clauses/sentences **421**
      - and embodied cognition **238–40**
      - garden-path sentences **223**, **224–5**, **309–10**
      - and implicit learning **557–9**
      - parsing **220**, **222–6**
      - syntax **220–2**, **241–2**
  - sentence verification **171**
  - Sergent, J. **147**
  - serial information capture **42**, **43**
  - serial order, and the phonological loop **332–5**
  - serial processing, of visual information **52**
  - serial recall **60**
  - serial search, in visual displays **55–6**
  - serial visual presentation, rapid **50–4**
  - Service, E. **330**
  - set effects, in problem solving **354**
  - SEU (subjective expected utility theory) **345**, **384–92**, **411**
    - axioms underlying **387–8**
    - and behavioural decision research **392–6**
    - and prospect theory **397**, **399**
      - violations of the axioms **388–91**, **392**
  - shadowing (dichotic listening) **43**, **61**
  - shadowing (head shadowing) **39–40**
  - Shaffer, W.O. **57**
  - Shafir, E. **390**, **395**
  - Shallice, T. **3**, **325**, **336**, **572**, **633**
  - shallow processing **240–2**
  - Shanks, D.R. **559**
  - Shannon, Claude **16**
  - Shiffrin, R.M. **270**, **309**, **545**, **560–2**
  - Shillcock, R. **212**
  - Shoben, E.J. **179**
  - short-term memory (STM) **307**
    - and consciousness **571**
    - and the irrelevant speech effect **320**
    - and long-term memory **308–10**, **330**
    - and reasoning **426**
    - and the structure of working memory **314**
    - see also* working memory
  - Siegel, R.S. **604**
  - similarity
    - and categorization **181–4**, **188–9**
    - and perceptual organization **78–9**
  - Simon, H.A. **25**, **356–7**, **366**, **367**, **369**, **411**, **413**
  - Singer, J.A. **524–5**
  - Singer, Jerome, and the Schachter-Singer theory of emotion **495**, **496–7**
  - SIUs *see* semantic information units (SIUs)
  - skill acquisition
    - arithmetic skills and the ACT-R model **601–8**
    - memories of **518**
    - and problem solving **374–5**
    - and the ventral and dorsal systems **105**

- skin conductance response (SCR) **144–5**  
 Skinner, B.F. 12, 16, 463  
*Verbal Behavior* 13  
 Sloman, S.A. 188–9  
 slots, and chunks in ACT-R architecture **587, 594–6**  
 Slovic, P. 391, 392, 393, 394  
 SME (structure mapping engine) 364–5  
 Smith, E.E. 176, 188–9  
 Smolensky, P. 582  
 Smyth, M.M. 316  
 SOA (stimulus onset asynchrony) **47, 50**  
 Soar architecture 613  
 social exchange, and the deontic selection task 447–8  
 somatic markers **480, 503**  
 sorting task **165–6**  
 sounds  
   attending to 41–2  
   disentangling 37–41  
 speech  
   dialogue 231, 245, 253–8  
   irrelevant speech effect 60–1, 320, 328  
   segmentation **198–202**  
   self-monitoring 251–3  
   speech errors and the language production system 245–9  
   waveforms 198–9, 633  
   *see also* spoken word recognition  
 Spellman, B.A. 363  
 Sperber, Dan 642  
 Sperling, G. 46–7, 49  
 spinal injury, and theories of emotion 495–6  
 spoken language  
   and dialogue 231  
   and Fodor's theory of modularity 636  
 spoken word recognition 198–206, 617, 618  
   cohort model of 159, 202–3  
   and Fodor's theory of modularity 636, 639  
   and lexical competition 204–6  
   and parallel activation **202–4**  
   segmenting the speech stream 198–202  
   TRACE model of 159, 201, 204–6, 215  
 spotlight of attention **41**  
 Squire, L.R. 282–3, 290, 299, 554  
 stage theory of cognitive development **418**  
 Stasz, C. 373  
 state emotion **482**  
 state-space, and problem solving 344, 346, 357, 359, 360  
 stereopsis, and Marr's theory of perception 95  
 Sternberg, R.J. 370  
 stimulus onset asynchrony (SOA) **47, 50**  
 STM *see* short-term memory (STM)  
 Stoljar, D. 646, 647  
 Stone, T. 645  
 Strack, F. 555  
 stranding speech errors 246–7, 248  
 Strevens, M. 187  
 Stroop effect 57, 58  
   Emotional Stroop task 58, 488–9  
 structural alignment, in problem solving 364  
 structural descriptions, and object recognition **121–2**  
 structure mapping engine (SME) 364–5  
 structure-building words **421**  
 sub-symbolic cognitive architectures **609–10**  
 subjective expected utility theory *see* SEU (subjective expected utility theory)  
 subjective self, and emotion 468  
 subliminal presentation 556  
   and Zajonc's theory of emotion 500  
 subordination, in language processing 241–2  
 subvocalization, and phonological working memory 318–19  
 supervisory attentional system (SAS) 325–6, 572  
 suppression effect, and everyday reasoning **432–7**  
 supraliminal presentation **556–8**  
 Sweller, J. 374  
 Swinney, D.A. 218, 635  
 symbol grounding problem **238**  
 symbol systems 24–5, 26  
 symbolic representations **610**  
 sympathetic ANS (autonomic nervous system) **465, 466**  
 synapses **501**  
 syntactic ambiguity **217, 224**  
 syntax 220–2, 241–2  
 systematicity, and connectionist modelling **631–2**  
 Taatgen, N.A. 609  
 Tanaka, J.W. 136  
 Tanenhaus, M.K. 225–6  
 TAP (transfer appropriate processing) **284–5**  
   and implicit memory 288–9, 291  
 target words **203**

- TAT (Thematic Apperception Test) 523  
 Taylor, A.M. 131  
 Teasdale, J.D. 485  
 template matching, and object recognition **120**, 122  
 temporal grouping effect **334–5**  
 text interpretation 233–40  
     and anaphora resolution 233–4  
     and inferences 236–7  
     knowledge, meaning and embodiment 237–40  
     and non-literal meaning 235–6  
     relating language to knowledge 237  
     and word meaning 234–5  
 texture gradient, and Gibson's theory of perception **84–6**  
 Thematic Apperception Test (TAT) 523  
 thematic role assignment **221**  
 theoretical issues 617–53  
     cognitive and neurobiological theories 461, 644–6  
     computation and cognition 619–32  
     and mental representations 617–18, 631  
     modularity 460, 632–43  
     radical neuron doctrine 646–8  
     *see also* CMM (computational model of the mind); connectionist models  
     ‘theory’-theory of concepts 158, 180–4, 185, 188, 190  
     thinking 1, 344–456  
         and symbol systems 25  
         *see also* problem solving  
 Thomson, N. 317  
 Thorndyke, P.W. 373  
 Thorne, A. 525  
 three-dimensional (3D) objects  
     perception of 73, 80, 91, 97  
     recognition of 119, 122, 124–35  
         3D object-centred descriptions **123**, 124  
         Biederman's theory 131–4  
         catalogue of 3D models **129–31**  
         Marr and Nishihara's theory 124–31, 132  
 threshold value, PDP models **580**, 581  
 Tolman, E.C. 14  
 top-down processing  
     and perception **77**, 98, 99, 105  
         visual masking 106–8  
     and visual word recognition 206–7  
 TOTE (Test-Operate-Test-Exit) units **15–16**  
 touch, object recognition by 118–20  
 Tower of Hanoi problem 356–8  
 Tower of London task 359–61  
 Towse, J.N. 324  
 TRACE model of spoken word recognition 159, 201, 204–6, 215  
 training, PDP models **581–2**  
 trait emotion **482**, 483  
 transfer appropriate processing *see* TAP  
 (transfer appropriate processing)  
 transitivity  
     and SEU (subjective expected utility theory) **387**, 388, 389–90  
     and categorization judgements 173–4  
 trauma memories *see* PTSD (post-traumatic stress disorder)  
 Traxler, M.J. 234–5  
 Treisman, A. 43–4, 49, 50, 55  
     feature integration theory **56–9**  
 Trueswell, J.C. 224  
 truth tables 421–2, 428–9, 434–5, 440  
 Tulving, E.  
     and autobiographical memory 520, 521, 522  
     and encoding specificity 284  
     and implicit memory 286, 287, 290  
     and modularity 633  
     and multiple memory systems 278, 279, 281, 299  
     processing levels in memory 271  
     and remember and know judgements 295, 296  
 Turing, A.M. 16  
 Turing test 17, 18  
 Turner, M.L. 313  
 Turner, T.J. 474  
 Turvey, M.T. 50  
 Tversky, A. 389–90, 391, 394, 395, 397, 399, 402, 404–5, 405, 407–8  
 two-dimensional (2D) objects  
     perception of 73, 81, 82  
     recognition of 120–2  
 Tyler, L.K. 222, 626–7, 628  
 Type I processing in memory **272–3**  
 Type II processing in memory **272**  
 typicality ratings of concepts  
     classical view of **170–2**  
     context-sensitivity of 179  
     prototype theory of 177–9  
     and similarity 180–1

- U-shaped pattern of learning, and cognitive modelling 608, 609  
 Ullman, M.T. 625  
 unconscious people, studying learning in 559–60  
 understanding production rules, and ACT-R models 589  
 uniqueness point, in spoken word recognition 203  
 units, in PDP models 580–1  
 unseen information, knowing about 46–9  
 utilization behaviour 325
- Vaidya, C.J. 291  
 valence 474, 476  
 Vallar, G. 321, 330–1  
 values, and chunks in ACT-R architecture 587, 594–6  
 Van Orden, G.C. 210  
 variables, and production rules in ACT-R models 589  
 varied mapping condition 561–2  
 vectors, and Marr’s theory of perception 95–6  
 ventral stream  
     and attention 64  
     and face recognition 145–7  
     and perception 102–4, 105, 105–6  
 ventriloquism effect 61  
 verbal labels for emotions 473–4  
 verbal protocols, solving 351, 352  
 verbal reasoning, effect of irrelevant memory load on 311  
 verbal report, and implicit learning 557  
 viewer-centred descriptions of objects 122–3  
     and Marr’s theory of perception 95–6, 123, 124  
 visual attention 45–54  
     and distraction 60, 61  
     and hearing 45–6  
     knowing about unseen information 46–9  
     and parallel processing 49  
     and Rapid Serial Visual Presentation (RSVP) 50–4  
 visual displays 55–9  
     ‘flanker’ effect 57–9  
     non-target effects 56–7  
     serial and parallel search in 55–6  
 visual illusions  
     and bottom-up perception 98–9  
     and perceptual errors 71–3
- visual masking *see* masking  
 visual pathways, and theories of perception 104–5  
 visual perception *see* perception  
 visual word recognition 206–13, 617, 619–20  
     DRC model of 208–9, 617, 618, 619–20, 629–30  
     and eye movements in reading 210–13  
     featural level of 206  
     IAC model of 206–7, 215  
     lexical effects in 207  
     mappings between spelling and sound 207–10  
 visuo-spatial working memory 315–16, 320  
 vocabulary acquisition, and the phonological loop 329–31  
 Vogel, E.K. 51, 52  
 Vollmeyer, R. 374  
 Von Eckardt, B. 24  
 von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behaviour* 384, 388
- Wade, N.J. 90  
 Wagenaar, W.A. 409  
 Wagner, J.L. 241  
 Walker, P. 316  
 Waller, A. 316  
 Warrington, E.K. 3, 131  
 Wason’s selection task 346, 437–48  
     abstract selection 437–42  
     deontic selection 443–8  
 water jars task 354–5, 359  
 Watson, J.B. 10, 12, 16  
 Watson, John B. 463  
 wavelengths 39–41  
 weight of links, PDP models 580, 581  
 Weiner, Norbert 15  
 well-calibrated judgements 409  
 well-defined categories 189  
 Wellman, H. 185, 187  
 Westenberg, C. 252  
 Wetherell, M. 165  
 Wheeler, M.A. 278  
 Wilson, B.A. 567–8  
 Winkler, R.L. 403  
 within-category distinctions  
     in face recognition 135–6  
     in object recognition 116, 134  
 witnesses, and face recognition 117

- Wittgenstein, L. 174  
 Woike, B. 523–4  
 Woldorff, M.G. 64  
 Wood, B. 44  
 word meanings  
     accessing 215–19  
     and text interpretation 234–5  
 word recognition 198–213, 641  
     *see also* spoken word recognition; visual word recognition  
 word-initial cohort 202–3  
 word-length effect, and phonological working memory 318  
 word-monitoring experiments, and information encapsulation 639–41  
 word-stem completion task 553  
 words, concepts and categories 164  
 words and rules model, and the past-tense debate 622–4  
 wordspotting task 200–1, 206  
 working backwards strategy, and novice problem solving 368  
 working forwards strategy, and expert problem solving 368  
 working memory 266–7, 307–41  
     and computer modelling 267  
     and consciousness 551, 560, 571–3, 574  
     and controlled processes of consciousness 546  
     and the dual-task paradigm 310–13  
     executive processes in 323–8, 336, 572  
     and Fodor’s theory of modularity 636  
     and individual differences 267  
     and problem solving 359  
     and protocol analysis 350, 351  
     structure of 314–29  
         multi-component model 314–17  
         phonological 317–23  
     and the ‘working self’ 459  
     *see also* phonological loop; short-term memory (STM)  
 working self 510–11, 517, 522–5, 529, 538  
     and childhood amnesia 512  
     and generative retrieval 525  
     and PTSD (post-traumatic stress disorder) 531–2, 534, 536, 537  
 Wright, M.J. 375  
 written language 232–44, 255  
     and coherence 232–3, 234  
     and cohesion 232, 233  
     and communication 231  
     and perspective in communication 240, 242–4  
     and shallow processing 240–2  
     *see also* text interpretation  
 WRUs (word recognition units) 141  
 Wundt school of introspection 8–9, 550  
 Würzburg school of introspection 9  
 X-ray problem 353–4, 363  
 Yerkes-Dodson law 478  
 Yin, R.K. 136, 149  
 Young, A.W. 138–9, 145  
     Bruce and Young model of face recognition 139–41, 617, 618, 619  
 Young, M.J. 97  
 Yule, G. 233  
 Zajonc, R.B. 556  
     and affective priming 565–6  
     theory of emotion 499–500, 500–1, 502  
 Zeki, S. 563, 573  
 Zhang, J. 358  
 Zuse, Conrad 16

# Acknowledgements

Grateful acknowledgement is made to the following sources for permission to reproduce material within this book:

## Cover image

Face icon based on image supplied by Getty Images.

## Colour plates

*Plates 1 and 2*: Tyler, L.K. and Moss, H.E. (2001) ‘Towards a distributed account of conceptual knowledge’, *Trends in Cognitive Science*, vol.5, no.6. Copyright © Elsevier Science Ltd; *Plates 6, 7 and 9*: Calder, A.J., Lawrence, A.D. and Young, A. W. (2001) ‘Neuropsychology of fear and loathing’, *Nature Reviews Neuroscience*, vol.2, no.5, May 2001, Nature Publishing Group.

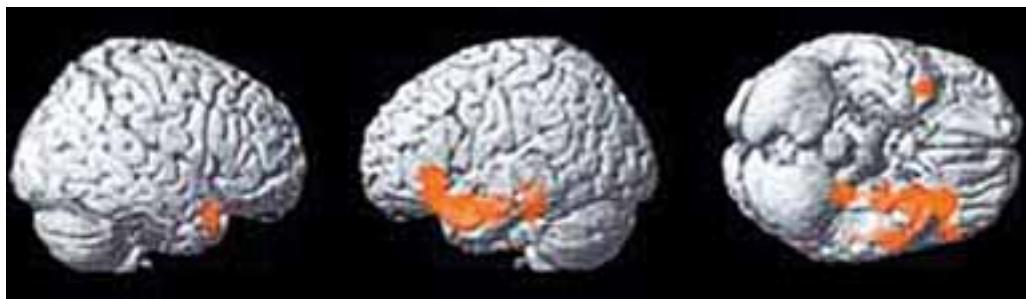
## Tables

*Table 5.4*: Reprinted from *Cognition*, vol.13, Armstrong, S.L. et al., ‘What some concepts might not be’, pages 263–308. Copyright (1983), with permission from Elsevier.

## Figures

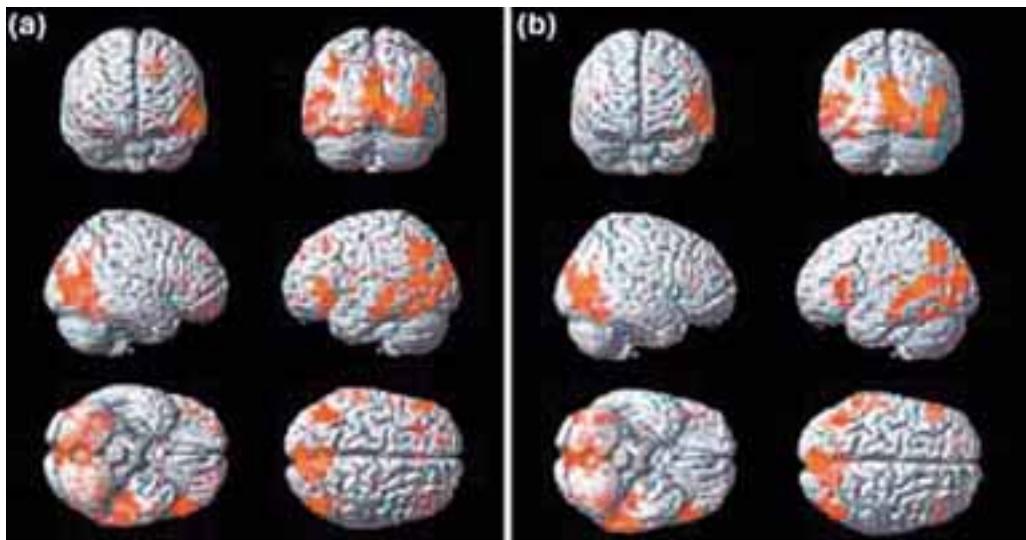
*Figure 3.7*: © ADAGP, Paris and DACS, London 2004/Bridgeman Art Library; *Figure 3.12*: © James J Gibson, *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates Inc., New Jersey, 1986, Fig. 2.1; *Figure 3.23*: © Nik Williams/The Swansea Museum 1996, with permission; *Figures 4.11, 4.14 and 4.15*: Marr, D. (2000) *Vision*, Henry Holt and Company, Inc.; *Figure 4.12*: © Succession Picasso/DACS 2004; *Figures 4.16, 4.17 and 4.18*: Marr, D. and Nishihara, H.K. (1978) ‘Representation and recognition of the spatial organization of three dimensional shapes’, *Proceedings of The Royal Society of London*, B200, pp.269–94. The Royal Society; *Figure 4.24*: Reprinted from *Trends in Cognitive Science*, vol.4, J.V. Haxby et al., ‘The distributed human neural system for face perception’, pages 223–33, Copyright (2000), with permission from Elsevier; *Figure 7.4*: Reprinted from the *Journal of Experimental Child Psychology*, vol.3, Krauss, R.M. and Weisberg, R.W., ‘Referential communication in nursery school’, page 33. Copyright © (1966), with permission from Elsevier; *Figure 9.1*: Atkinson, R.C. and Shiffrin, R.M. (1971) ‘The control of short-term memory’, *Scientific American*, vol.225. By permission of Scientific American; *Figure 13.1*: ‘Automatic Nervous System’ figure from *The Penguin Dictionary of Psychology* by Arthur S. Reber (Penguin Books 1985, third edition 2001). Copyright © Arthur C. Reber, 1985, 1995, 2001; *Figure 13.3a*: Ekman, P. and Friesen, W.V. (1975) *Unmasking the Face*. Copyright © Paul Ekman; *Figure 13.4, top left*: © J & P Wegner/Foto Natura/FLPA, *top right*: © Jan Pitman/Associated Press, *bottom right*: © Photodisc Europe; *Figure 13.13*: Adapted from Rosenzweig, M.R. et al. (1999) *Biological Psychology: An Introduction to Behavioural, Cognitive, and Clinical Neuroscience*, Sinauer Associates; *Figure 15.4*: Baddeley, A. and Wilson, B.A. (1994) ‘When implicit learning fails: amnesia and the problem of error elimination’, *Neuropsychologia*, vol.32, no.1. Elsevier Science Ltd; *Figures 16.6, 16.9, 16.17 and 16.20*: Adapted from Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Lawrence Erlbaum Associates Inc.; *Figure 17.1*: Reprinted from *Trends in Cognitive Science*, vol.6, no.11, Pinker, S. and Ullman, M.T., ‘The past and future of the past tense’, page 457, Copyright © (2002), with permission from Elsevier.

Every effort has been made to contact copyright owners. If any have been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.



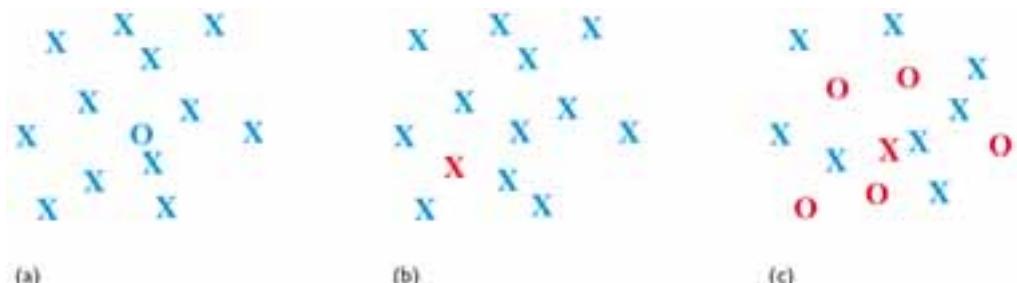
**Plate 1** Results of PET studies using written words in lexical decision and semantic categorization tasks. Red areas show activated regions. No region showed a significant difference in activation between living and non-living things

Source: Tyler and Moss, 2001, Figure 1, p.247



**Plate 2** Results of fMRI study using pictures of living and non-living things in a semantic categorization task. The active brain regions are areas associated with non-living things (a), and with living things (b). No regions showed significant differential activation for living and non-living things

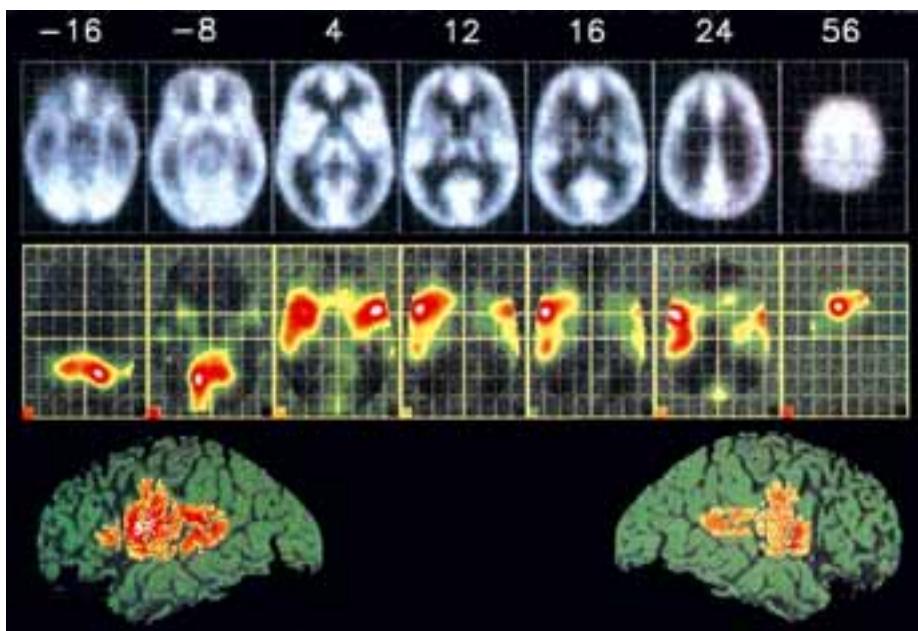
Source: Tyler and Moss, 2001, Figure 2



**Plate 3** Typical stimuli used in Triesman's experiments. Find the odd item in each of the groups, (a), (b) and (c)

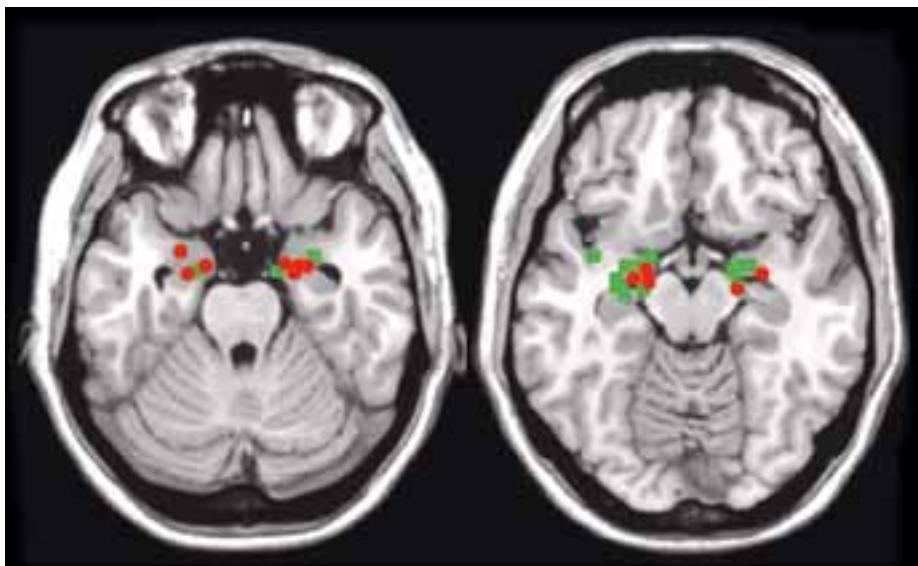
Cat	Small	Red	Black
Door	Window	Green	Red
Chair	Cat	Blue	Red
Small	Sleep	Black	Purple
Window	Chair	Red	Blue
Great	Great	Yellow	Yellow
Sleep	Door	Purple	Green
Long	Long	Grey	Grey
Chair	Cat	Blue	Red
Great	Small	Yellow	Black
Sleep	Chair	Purple	Blue
Cat	Great	Red	Yellow
Window	Sleep	Pink	Purple
Long	Door	Grey	Green

**Plate 4** The Stroop effect. The task is, *as quickly as possible*, to name the colour in which each word is printed: do *not* read the words. Try the first two columns to start with – the task is not too difficult. Then try the second two columns. The conflicting colour words are very disruptive



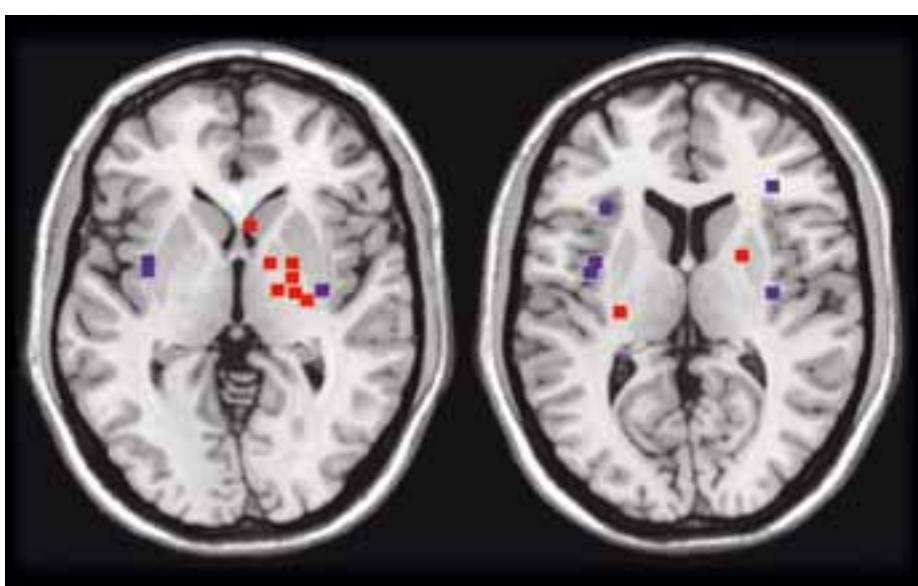
**Plate 5** Brain activity associated with phonological processing. The second row shows brain activity for both phonological and control tasks combined; activity is shown for different horizontal ‘slices’ through the brain taken at different points, and indexed by the numbers in the first row. The third row shows increases in activity associated with phonological processing, reflecting the difference between phonological and control tasks. In the fourth row, these areas of increased activity have been mapped onto the cortex of the brain

Source: Paulescu et al., 1993, Figure 2, p.343



**Plate 6** Amygdala activations in response to the processing of fearful faces (green squares) or learning about fear (red circles). The image on the left is a horizontal 'slice' through the brain, with the eyes at the top end and the back of the head at the bottom. The image on the right is the same sort of slice, but taken higher up, more towards the top of the head. There is a tendency for the activation triggered by processing fearful faces to involve the left amygdala, whereas learning about fear seems to produce more bilateral activation

Source: Calder et al., 2001, Figure 3, p.357



**Plate 7** Insula and basal ganglia activations in response to disgust. The two images depict different slices through the brain. The insula activations are shown in purple and basal ganglia activations in red. The basal ganglia signals are mainly in the right hemisphere, whereas the insula signals are more evenly distributed across the two hemispheres

Source: Calder et al., 2001, Figure 4, p.358

	Standard Stroop task		Emotional Stroop task	
trial 1	BLUE ↓ RED	GREEN	CANCER	HOUSE
trial 2		BROWN	DANGER	LAUGH
trial 3	BROWN ↓ GREEN ↓ RED	compared with BLUE RED BROWN	ATTACK TUMOUR HORROR DEATH REVENGE EVIL	ANIMAL MODERN PICTURE FATHER BEAUTY COOK
etc.			Emotion condition	Neutral condition
	Incongruent condition	Congruent condition		

Task: 'Name the ink colour as fast as possible'

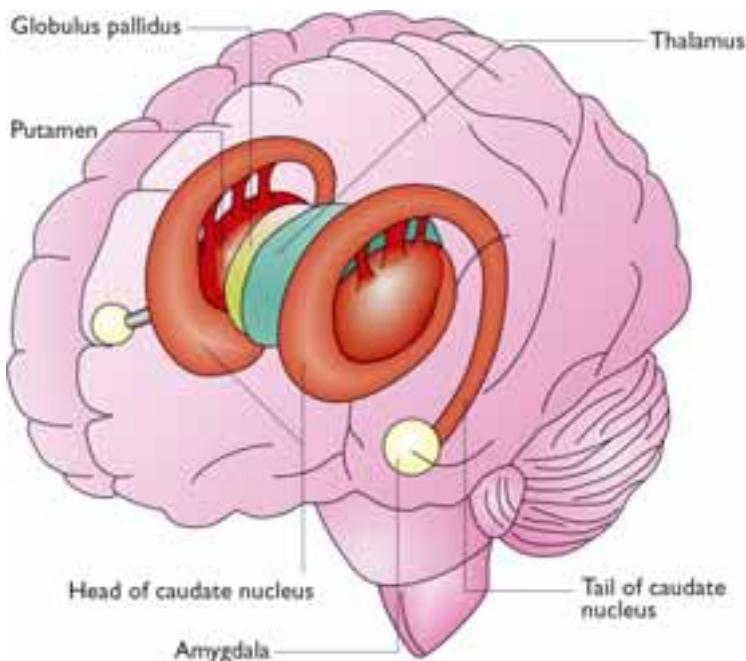
#### Standard result

Incongruent slowed compared with congruent condition (because word meaning interferes with colour naming on incongruent trials).

#### Emotional result

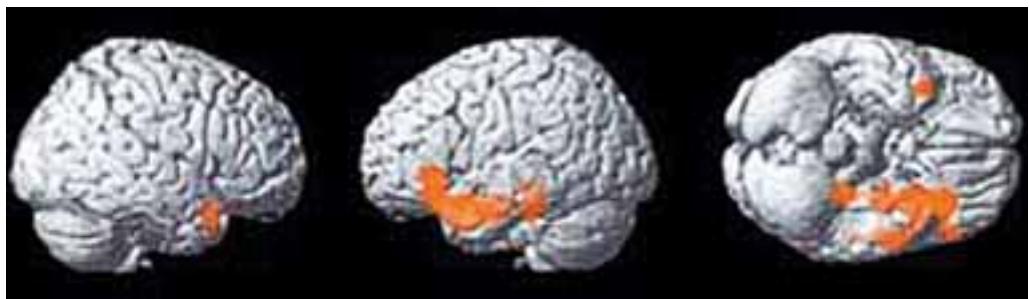
Anxious participants: emotion condition slowed compared with neutral condition.  
Non-anxious participants: no difference between emotion condition and neutral condition.

**Plate 8** The standard and emotional Stroop



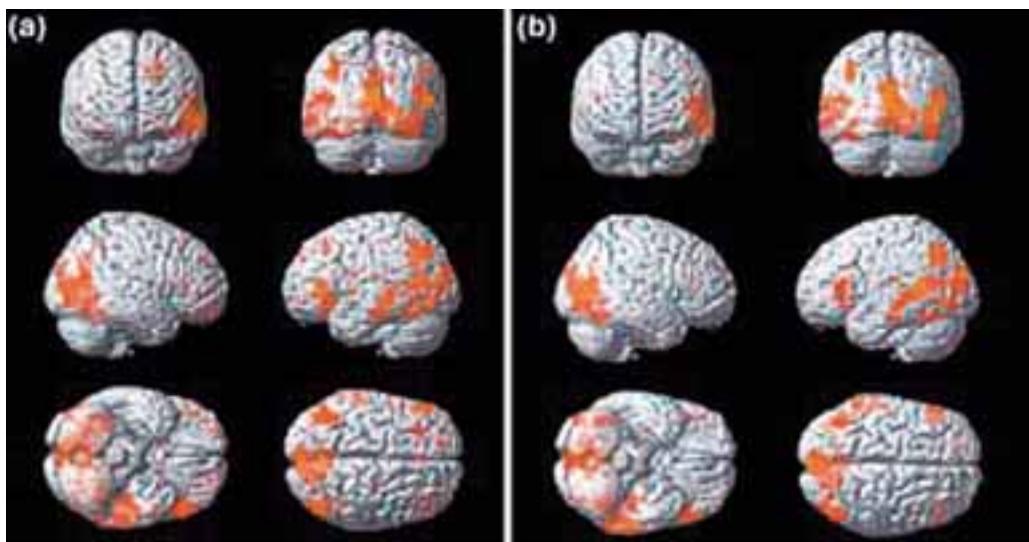
**Plate 9** A schematic drawing of the human brain showing the thalamus and amygdala

Source: Calder et al., 2001, Figure 1(a), p.353



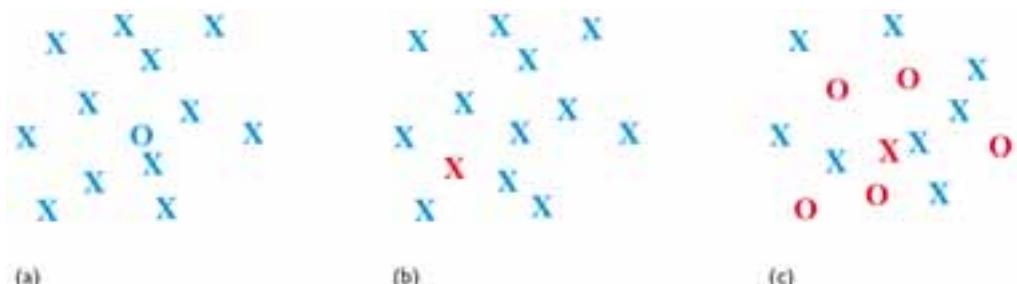
**Plate 1** Results of PET studies using written words in lexical decision and semantic categorization tasks. Red areas show activated regions. No region showed a significant difference in activation between living and non-living things

Source: Tyler and Moss, 2001, Figure 1, p.247



**Plate 2** Results of fMRI study using pictures of living and non-living things in a semantic categorization task. The active brain regions are areas associated with non-living things (a), and with living things (b). No regions showed significant differential activation for living and non-living things

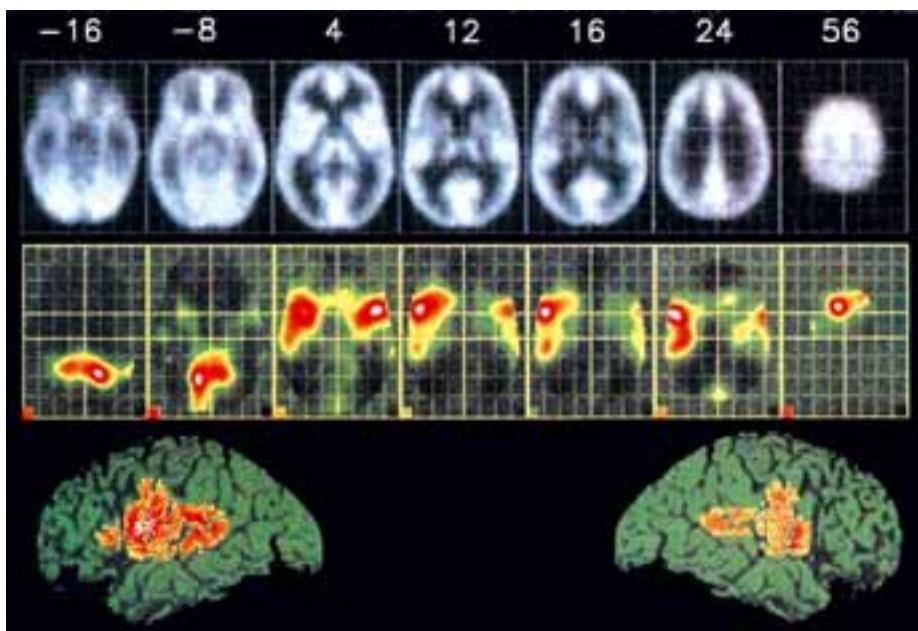
Source: Tyler and Moss, 2001, Figure 2



**Plate 3** Typical stimuli used in Triesman's experiments. Find the odd item in each of the groups, (a), (b) and (c)

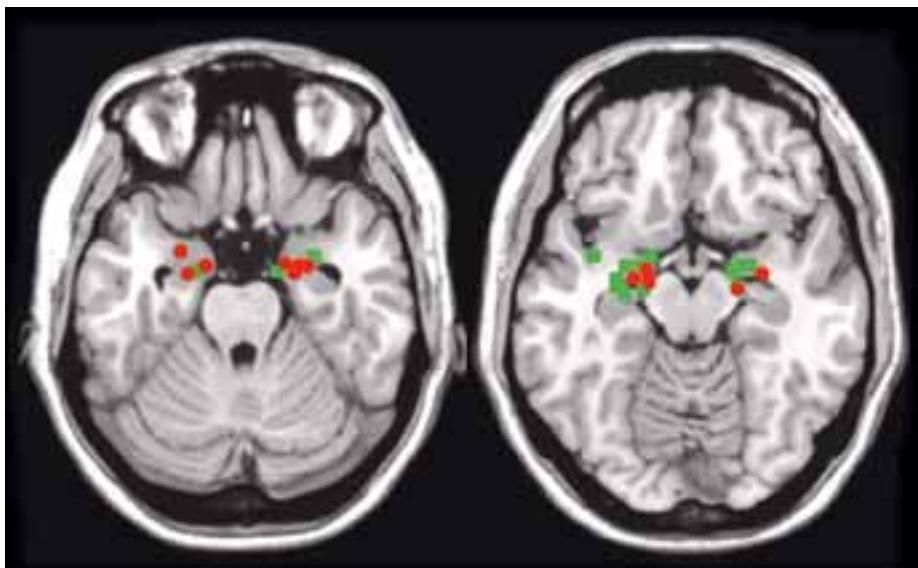
Cat	Small	Red	Black
Door	Window	Green	Red
Chair	Cat	Blue	Red
Small	Sleep	Black	Purple
Window	Chair	Red	Blue
Great	Great	Yellow	Yellow
Sleep	Door	Purple	Green
Long	Long	Grey	Grey
Chair	Cat	Blue	Red
Great	Small	Yellow	Black
Sleep	Chair	Purple	Blue
Cat	Great	Red	Yellow
Window	Sleep	Pink	Purple
Long	Door	Grey	Green

**Plate 4** The Stroop effect. The task is, *as quickly as possible*, to name the colour in which each word is printed: do *not* read the words. Try the first two columns to start with – the task is not too difficult. Then try the second two columns. The conflicting colour words are very disruptive



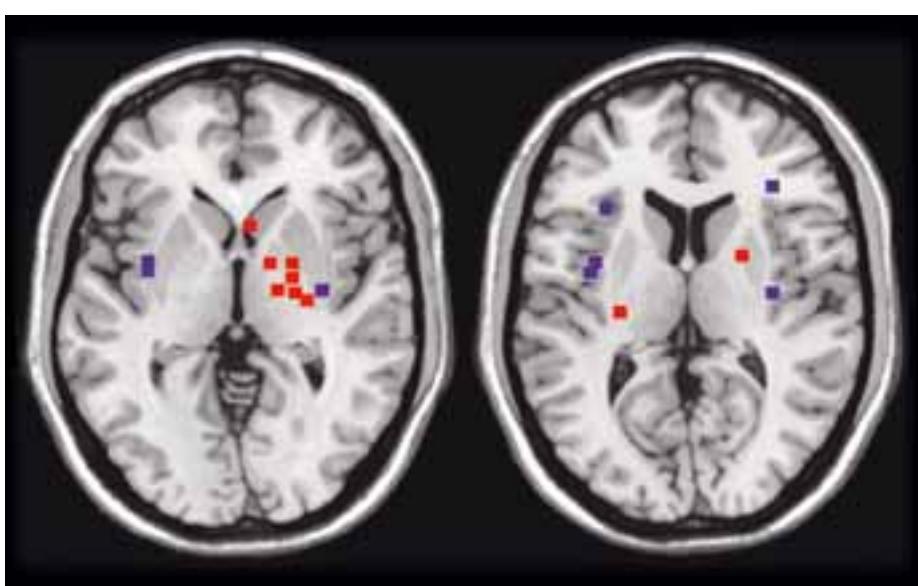
**Plate 5** Brain activity associated with phonological processing. The second row shows brain activity for both phonological and control tasks combined; activity is shown for different horizontal ‘slices’ through the brain taken at different points, and indexed by the numbers in the first row. The third row shows increases in activity associated with phonological processing, reflecting the difference between phonological and control tasks. In the fourth row, these areas of increased activity have been mapped onto the cortex of the brain

Source: Paulescu et al., 1993, Figure 2, p.343



**Plate 6** Amygdala activations in response to the processing of fearful faces (green squares) or learning about fear (red circles). The image on the left is a horizontal 'slice' through the brain, with the eyes at the top end and the back of the head at the bottom. The image on the right is the same sort of slice, but taken higher up, more towards the top of the head. There is a tendency for the activation triggered by processing fearful faces to involve the left amygdala, whereas learning about fear seems to produce more bilateral activation

Source: Calder et al., 2001, Figure 3, p.357



**Plate 7** Insula and basal ganglia activations in response to disgust. The two images depict different slices through the brain. The insula activations are shown in purple and basal ganglia activations in red. The basal ganglia signals are mainly in the right hemisphere, whereas the insula signals are more evenly distributed across the two hemispheres

Source: Calder et al., 2001, Figure 4, p.358

	Standard Stroop task		Emotional Stroop task	
trial 1	BLUE ↓ RED	GREEN	CANCER	HOUSE
trial 2		BROWN	DANGER	LAUGH
trial 3	BROWN ↓ GREEN ↓ RED	compared with BLUE RED BROWN	ATTACK TUMOUR HORROR DEATH REVENGE EVIL	ANIMAL MODERN PICTURE FATHER BEAUTY COOK
etc.			Emotion condition	Neutral condition
	Incongruent condition	Congruent condition		

Task: 'Name the ink colour as fast as possible'

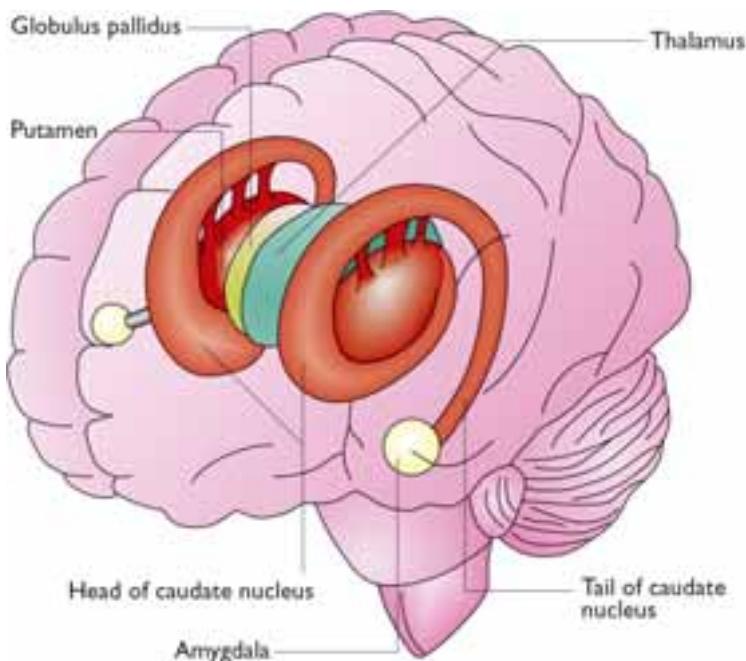
#### Standard result

Incongruent slowed compared with congruent condition (because word meaning interferes with colour naming on incongruent trials).

#### Emotional result

Anxious participants: emotion condition slowed compared with neutral condition.  
Non-anxious participants: no difference between emotion condition and neutral condition.

**Plate 8** The standard and emotional Stroop



**Plate 9** A schematic drawing of the human brain showing the thalamus and amygdala

Source: Calder et al., 2001, Figure 1(a), p.353