Knowledge and Language

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

Editors

ROBERT S. COHEN, *Boston University*
JÜRGEN RENN, *Max-Planck-Institute for the History of Science*
KOSTAS GAVROGLU, *University of Athens*

Editorial Advisory Board

THOMAS F. GLICK, *Boston University*
ADOLF GRÜNBAUM, *University of Pittsburgh*
SYLVAN S. SCHWEBER, *Brandeis University*
JOHN J. STACHEL, *Boston University*
MARX W. WARTOFSKY†, *(Editor 1960–1997)*

VOLUME 227

KNOWLEDGE AND LANGUAGE

Selected Essays of L. Jonathan Cohen

by

L. JONATHAN COHEN
Queen's College, Oxford,
and *Fellow of the British Academy, United Kingdom*

edited and with an introduction by

JAMES LOGUE
Somerville College,
Oxford, United Kingdom

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-5955-0 ISBN 978-94-017-2020-5 (eBook)
DOI 10.1007/978-94-017-2020-5

Printed on acid-free paper

All Rights Reserved
© 2002 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 2002
No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

Contents

Preface vii

Publisher's Note ix

Introduction xi
James Logue

1
On the Project of A Universal Character 1

2
On A Concept Of Degree Of Grammaticalness 15

3
The semantics of metaphor 27

4
Can The Logic Of Indirect Discourse Be Formalised? 41

5
Some Remarks On Grice's Views About The Logical Particles Of Natural Language 49

6
Can The Conversationalist Hypothesis Be Defended? 67

7
How Is Conceptual Innovation Possible? 77

8
Natural Language Definitions 95

9
A Problem About Ambiguity In Truth-Theoretical Semantics 113

10
The Individuation of Proper Names 119

11
Third World Epistemology 141

12
Guessing 155

13
Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses 175

14
Are People Programmed to Commit Fallacies? 195

15
Inductive Logic 1945-1977 221

16
Some Historical Remarks On the Baconian Conception of Probability 245

17
Twelve Questions about Keynes's Concept of Weight 261

18
Some Steps towards a General Theory of Relevance 279

19
Should A Jury Say What It Believes Or What It Accepts? 293

20
Are There Ethical Reasons For Being, Or Not Being, A Scientific Realist? 313

Preface

I am very grateful to Kluwer Academic Publishers for the opportunity to republish these articles about knowledge and language.

The *Introduction* to the volume has been written by James Logue, and I need to pay a very sincerely intended tribute to the care and professionalism which he has devoted to every feature of its production.

My thanks are also due to Matthew McGrattan for his technical assistance in scanning the articles onto disk and formatting them.

L. Jonathan Cohen

Publisher's Note

Thanks are due to the following publishers for permission to reproduce the articles in this volume.

Paper 1	On the project of a universal character. Oxford University Press.
Paper 2	On a concept of a degree of grammaticalness. Logique et Analyse.
Paper 3	The semantics of metaphor. Cambridge University Press.
Paper 4	Can the logic of indirect discourse be formalised? The Association for Symbolic Logic.
Paper 5	Some remarks on Grice's views about the logical particles of natural language. Kluwer Academic Publishers.
Paper 6	Can the conversationalist hypothesis be defended? Kluwer Academic Publishers.
Paper 7	How is conceptual innovation possible? Kluwer Academic Publishers.
Paper 8	Should natural language definitions be insulated from, or interactive with, one another in sentence composition? Kluwer Academic Publishers.
Paper 9	A problem about truth-functional semantics. Basil Blackwell Publisher Ltd.
Paper 10	The individuation of proper names. Oxford University Press.
Paper 11	Some comments on third world epistemology. Oxford University Press.
Paper 12	Guessing. The Aristotelian Society.
Paper 13	Bayesianism versus Baconianism in the evaluation of medical diagnoses. Oxford University Press.
Paper 14	Are people programmed to commit fallacies? Basil Blackwell Publishers, Ltd.
Paper 15	Inductive logic 1945-1977. Kluwer Academic Publishers.
Paper 16	Some historical remarks on the Baconian conception of probability. The Journal of the History of Ideas.
Paper 17	Twelve questions about Keynes's concept of weight. Oxford University Press.
Paper 18	Some steps towards a general theory of relevance. Kluwer Academic Publishers.
Paper 19	Should a jury say what it believes or what it accepts? The Cardozo Law Review.
Paper 20	Are there ethical reasons for being, or not being, a scientific realist? Kluwer Academic Publishers.

Introduction

James Logue

It is a privilege and a pleasure to be asked to write an introduction to this collection of papers by my former teacher and colleague L. Jonathan Cohen. Jonathan Cohen's career to date as a professional philosopher has spanned more than fifty years. During this time he has produced a very large body of extremely influential and important work – indeed, a great deal of it seminal and groundbreaking – in philosophy of science, the philosophy of induction and probability, epistemology, philosophical logic, the philosophy of language, jurisprudence and political philosophy, among other fields. The papers in philosophy of language and epistemology which make up this volume exhibit all the qualities most characteristic of Cohen's work: originality and insight, a distinctive analytical perspective, great clarity, rigour and cogency of argument, elegance and wit. Their reproduction here will, I hope, be a stimulus to new work on a wide range of important topics.

With the obvious exception of the previous paragraph, I have written this introduction in consultation with Jonathan Cohen: I hope that that has saved me from misrepresenting his views. But, of course, it should not be taken as a statement of his own overview of the papers. As to how it reflects my own views, it would obviously be inappropriate for an introduction of this sort to stray into becoming a critique: so, in summarising the papers, I have resisted the strong temptation to add comments of my own, except in a couple of clearly signalled instances.

It would be a mistake to see these twenty papers, written over a period of more than forty years, as part of a single consciously unified project. Nevertheless, there is a great deal of continuity and thematic unity as the papers move through philosophy of language via general epistemology to confirmation, induction and probability. It is, perhaps, easier to see this continuity by tracing a route backwards through the papers (not chronologically, but as they appear here). The last group of papers is largely

concerned with how empirical knowledge may be acquired through evidence in states of uncertainty. The middle group explores how such evidence often requires or results in conceptual innovation and is given to us in language the meaning of which is not always easy to determine. The first group asks how a theory of meaning can be constructed both for natural and for artificial languages. {That is, so to speak, the logical ordering of the thought behind each paper: the present order has been chosen to facilitate thinking one's way into the issues.} This thematic unity is underpinned by Cohen's common commitment to the richness both of language and of enquiry and his pluralist opposition to simplistic and dogmatic formalizations such as are offered by certain kinds of compositional semantics and by monistic interpretations of probability. In the remainder of this introduction, where I summarise each paper in turn, I shall regularly draw attention to the many connections among them which are generated by these threads which run through much of Cohen's philosophy.

The first ten papers focus on issues in the philosophy of language.

Paper 1 (*On the Project of a Universal Character* [1954]) begins by attacking a widely held view that Leibniz was the solitary originator in the seventeenth century of the project of developing a 'universal character' – an artificial language, suitable for the new science of the day. Leibniz intended his universal character to fill three main roles. It was to be an auxiliary language facilitating communication between scientists of different nations; it was to act as a simplified notation for science; and it was to be an instrument for discovery and demonstration. But a number of Leibniz's contemporaries and predecessors also articulated and attempted to carry out just such a project. Bacon [1605, 1623] and Descartes [1629] both discussed its possibility and desirability; from Descartes' time onwards the project seems to have become a commonplace, with numerous writers – the best known being Mersenne, Dalgarno and Wilkins – carrying it through in some detail and with a clear commitment to the need for a universal character to fill the three roles Leibniz identifies. What was distinctive about Leibniz's position was solely his conception of a calculus (independently important as a progenitor of modern formal logic) which would operate on the formulae of his character so as to lay bare the implications of what was already known. Cohen argues that the failure of these seventeenth-century projects - which contributed nothing towards either enhancing international communication or promoting scientific discovery – has implications for how we now ought to view the relationship between science and language. It is a fundamental mistake, he says [p. 11] "to think that the same language could serve adequately both as an unspecialised international

auxiliary and also as a scientific terminology". Unspecialised languages need a range and flexibility of meaning – ambiguities and all – which would be inappropriate to a specialised scientific notation. Thus, we should avoid any temptation to suppose that there is a set of compatible ideals in the light of which we could and should generate a 'perfect language'.

Paper 2 (*On a Concept of Degree of Grammaticalness* [1965]) has two main aims: first, to distinguish degree of grammaticalness (a concept emerging from discussions of generative grammar) from relative intelligibility; secondly, to explain the former by use of a theory of evidential support which accounts for the non-statistical character of generative grammar.

In *Syntactic Structures* [1957] Chomsky compared sentences like "John enjoyed the play and my friend liked it" to sentences like "John enjoyed and my friend liked the play". The latter could hardly be termed wholly non-grammatical, but in speech it shares numerous features with the reading aloud of wholly non-grammatical strings of words. This suggests that there is a sense in which the former sentence is more grammatical than the latter. Chomsky attempted to characterise degree of grammaticalness in terms of a hierarchy of the categories appearing in the rules of a generative grammar. Katz [1964] attacked these efforts at a purely syntactic theory of partial grammaticalness, arguing that a semantic component is also required which will account for the typical ambiguity of semi-sentences (those strings of a language L not generated by an optimal grammar of L but with sufficient structure to be understood by speakers of L). Cohen argues that Katz's view lacks all bite unless it identifies intelligibility with degree of grammaticalness. But there are simple and powerful counterexamples which show any such identification to be strongly counterintuitive. Instead, what is needed is a scale of intelligibility independent of non-semantic scales of grammaticalness.

How then should the degree of grammaticalness of a string in L be calibrated? Cohen suggests that it may be equated with the highest degree of evidential support obtainable for the statement that this string is fully grammatical. A small number of intuitively plausible assumptions show, he argues, that degree of evidential support in this context cannot be taken to be mathematical probability as standardly conceived. Instead, it is necessary, and possible, to define confirmation-functors for generative grammars which have different properties from mathematical probabilities: they turn out to have what Cohen later called a 'Baconian' structure (see the last third of the papers in this collection, as well as his books *The Implications of Induction* [1970] and *The Probable and the Provable* [1977]).

Paper 3 (*The Semantics of Metaphor* [1979]) argues that the nature of metaphor is a linguistic problem which requires clarification via the study of sentence-meaning rather than the study solely of speech-acts. It can be seen to be a linguistic problem by noting that it is characteristic of natural language sentences like "He is a lion" which can be read metaphorically as well as literally that they are indefinitely rich in possible meanings: whereas most artificial languages must lack any possibility of metaphor. However, it is not a problem which can be solved purely within a theory of speech-acts. Standard kinds of speech act such as promising, apologizing and so on are overridden when one moves from direct to indirect speech – when I report your apology I am not myself apologizing. But if I report that you said that he is a lion the metaphorical content is not overridden in this way: my report can be understood only by someone who understands the metaphor. So, "metaphorical meaning inheres in sentence-meaning, not just in speech acts." (p. 29)

One possible account of metaphors would then be to suggest that a term used metaphorically simply has a different meaning from that in its literal use: metaphor reduces to polysemy. Cohen rejects this "method of multiplication" in favour of a "method of cancellation". These rival methods may be applied to nonmetaphorical cases too. For instance, in most uses of the term "insult", a particular kind of intention is presupposed. If we take it that the sentence I: "It was an unintentional insult" is not self-contradictory, the method of multiplication claims that the term "insult" here is a different lexeme from the term as standardly used. The method of cancellation, Cohen argues, ought to be preferred on the grounds of lexical parsimony: it claims that in sentence I the feature + INTENTIONAL, normally part of the sense of "insult", is cancelled by the adjective "unintentional" but we have no need to suppose two distinct lexemes.

In earlier work (Cohen & Margalit [1972]) Cohen claimed that metaphor likewise involves feature cancellation, distinguished from the nonmetaphorical cases by always being the cancellation of semantically important features. He continues to insist (contrary to, for instance, Wittgensteinian family-resemblance theories of meaning) that an adequate semantic theory for natural language must assume some kind of importance-ordering for semantic features. But he no longer adheres to the [1972] basis for distinguishing metaphorical from nonmetaphorical feature-cancellations: he instead suggests making use of a topic-comment distinction within sentences in which cancellation occurs to yield an analysis of most metaphorical uses of expressions. Even then, though, the richness and variety of the deployment of metaphors in nat-

ural language suggests that it will be a complex and difficult task to construct a rigorous and adequate theory of metaphor.

Paper 4 (*Can the Logic of Indirect Discourse be Formalised?* [1957]) considers attempts over the preceding ten years by Carnap and Church to provide formalizations which will exhibit the logical structure of everyday utterances. In these formalizations, semantic antinomies arising from indirect discourse are supposed to be avoided by using a hierarchy of language-levels. But this method of analysis seems to lead to problems in the treatment of sentences such as P:

> "If the policeman testifies that anything, which the prisoner deposes, is false, and the prisoner deposes that something, which the policeman testifies, is true, then something, which the policeman testifies, is false and something, which the prisoner deposes, is true."

There are good reasons for accepting that P is a non-trivial logical truth. But whether we adopt, modify or abandon a hierarchy-based analysis, comprehensive formalization of statements like P appears to be extremely difficult, if not impossible.

Cohen argues that this is not merely a technical difficulty which a more sophisticated formalization may be expected to resolve. We may not often utter sentences like P. But in judicial or historical investigations, as well as in everyday discourse, we often incorporate, as P does, statements about a witness's truthfulness alongside the report of his testimony, as premisses in an argument about the facts which seems to require us to jump from one level of statement to another. It is such arguments, Cohen claims, not the natural languages in which they are expressed, which constitute the proper objects of formalization. Once this is accepted, the idea, or ideal, of a single hierarchy of formalized languages in which one could articulate every logical truth that can be uttered in a natural language ceases to hold any appeal.

Paper 5 (*Some Remarks on Grice's Views about the Logical Particles of Natural Language* [1971]) and paper 6 (*Can the Conversationalist Hypothesis be Defended?* [1979]) attack the hypothesis, which Grice broadly endorsed, that the familiar logical particles of natural language ('not', 'and', 'or' and 'if ... then ...') do not diverge in meaning from the standard interpretations of the corresponding formal-logical symbols ($\sim$, $\wedge$, $\vee$ and $\rightarrow$). On this view, which Cohen calls the Conversationalist Hypothesis, apparent divergences are due to pragmatically grounded standing presumptions about conversational utterances which generate conversational and conventional implicatures.

Cohen argues in paper 5 that we should instead prefer what he calls the Semantical Hypothesis: that some occurrences of these particles do diverge in meaning from the corresponding symbols, while some do not, and both cases should be explained within a semantic theory of

language. While not denying that proper regard should be paid to conversational presumptions, he opposes any replacement of the semantic by the pragmatic. For each particle in turn, he argues that its meaning or linguistic function is richer than that of its formal-logical counterpart, so that genuine divergence in meaning frequently occurs: however, there are statements where additional clauses cancel out these richer features (rather as discussed in paper 3) so that there is no such divergence. His central argument is that the Conversationalist Hypothesis, in rejecting divergence of meaning between the particles and their symbolic counterparts, is committed to the truth-functionality of the particles: 'and' must be purely truth-functional in its sense, exactly like '$\wedge$'. This may appear plausible when one considers only simple examples: but it can be shown to be clearly indefensible by examining more complex sentences, especially those where negation, conjunction or disjunction occurs embedded within the antecedent of a conditional.

The argument here echoes that made against speech-act theories of metaphor in paper 3.

Paper 6 responds to attempts by Ralph Walker [1975] to defend the Conversationalist Hypothesis against these criticisms. Walker claims that if the Conversationalist Hypothesis had such a strong commitment to this interpretation of the truth-functionality of 'and', 'or' etc. then it would be wildly implausible. Rather, it is committed only to the truth-functionality of the propositional content of embedded sentences, though what they convey may not be truth-functional at all: the truth-value of what is conveyed will depend on the conversational implicatures of their utterance. Cohen's reply is that close attention to Grice's work shows that he, at least, endorsed the stronger commitment, whether implausible or not. Walker is, then, in fact advancing a different hypothesis: but, Cohen argues, it is one which sacrifices much of what is initially appealing in Grice's position, generates various paradoxes and oddities, and is difficult to articulate without circularity.

Paper 7 (*How is Conceptual Innovation Possible?* [1986]) addresses this question philosophically, rather than historically or sociologically. It asks how the possibility of conceptual innovation, occurring as scientific knowledge grows, bears on the way in which we should theorize about language, meaning and how new word-meanings enter natural language. This paper thus very directly connects the twin concerns, with the nature of knowledge and of language, and the fundamental questions of how these interrelate, which underlie all the papers in this collection.

Cohen begins by distinguishing between two accounts of how the meaning of a sentence depends on the meaning of its constituent words. On an insulationist account, the meaning of each word is insulated

against interference from any other word in the sentence. An interactionist account denies this, holding that in some sentences the meaning of a word may be partly determined by its context in that sentence (though this is not always so, and the influence on word-meaning of other words in the sentence may be very slight). Generally, artificial languages tend to satisfy an insulationist account, natural languages an interactionist one. Artificial languages are, indeed, usually designed to fit the insulationist model (see paper 1): while the interactionist nature of natural languages results in their much greater expressive potential, as regards both literal and metaphorical meaning. Of course, the insulationist model will fit, to a very close approximation, the patterns of composition and resultant meaning of many sentences in natural languages. But it cannot be made to fit them all. Hence, Cohen claims, projects (such as those of Davidson and Montague) which aim to construct a compositional semantics for natural language along the same lines as formal systems cannot succeed.

How, then, on the interactionist view, is conceptual innovation possible? Cohen suggests that new word-meanings which are not generated simply by stipulative definition emerge from the unactualised interactionist potential of the language. These word-meanings have been latent in the language until – usually as the result of some novel theory or advance in knowledge – they occur for the first time in an actual utterance. Once that has happened, they may well become part of our customary usage and alter our common conceptual apparatus. This process is central to the development of science, Cohen concludes: if the project of a universal character (paper 1) had been successful, it would have – directly contrary to its advocates' intentions – stopped rather than supported scientific theorizing.

Papers 8 (*Should Natural-Language Definitions be Insulated from, or Interactive with, One Another in Sentence Composition?* [1993]) and 9 (*A Problem about Ambiguity in Truth-Theoretical Sentences* [1985]) follow on from Paper 7 in criticising interactionism. Paper 8 asks: "when a particular set of lexical definitions, in a particular syntactic framework, is supposed to compose the meaning of a given sentence, should those definitions be supposed to operate independently of one another, or not?" [p. 95]. In answering this question in the negative, the paper expands and develops the distinction between interactionism and insulationism, stressing that it properly belongs to semantic theory, not to psychology or pragmatics. Cohen makes it clear that his objection is to the simplistic dogmatism of the insulationist: he does not advocate an equally dogmatic universal interactionism.

In parallel to the account of probability offered in the later papers in this volume, he advocates a motivated pluralism (distinct from what he sees as a Wittgensteinian anomalism): he argues that insulationism is not appropriate everywhere, but nor is interactionism. Given an interactionist model, with insulationism on occasions a reasonable approximation, one can then see that an account of sentence-meaning will require not only definitions of the fundamental default meanings of the component words but also a range of semantic principles governing extended and modified word-meanings as they are influenced by their context in sentences. These principles are what enable users of natural languages, by contrast with artificial ones, to do without an implausibly large number of definitions while saying many different things.

Paper 9 focuses on the problems which verbal ambiguity poses for an insulationist truth-theoretical semantics. Syntactic ambiguity can be handled quite easily by such theories as Davidson's. But verbal ambiguity cannot be treated as a rare and deviant occurrence in natural languages, nor reduced to the existence of homonyms. It is a fundamental feature of natural languages, yet attempts to translate ambiguous sentences into disjunctive T-sentences produce T-sentences of infinite length.

Paper 10 (*The Individualisation of Proper Names* [1980]) continues the theme of resisting formalist reductions of natural language: it resists the temptation to suppose that proper names contrast sharply with predicates and that this contrast arises from their invariably designating just one object. Instead, Cohen advocates what he calls the 'linguistic' conception of a proper name as individuated by its phonological form alone, rather than the 'idiosyncratic' conception according to which proper names are individuated by their form-and-bearer. {Since this paper appeared in a *Festschrift* for Sir Peter Strawson, I cannot resist updating Cohen's illustrative example with a real-life one which could not have been deployed twenty years ago. On the linguistic view, 'Strawson' is a different proper name from 'Dummett', but the name of the author of *The Bounds of Sense* is, unequivocally, the name of the author of *The Secret Connexion.* On the idiosyncratic view, either the father's name is different from the son's or the form 'Strawson 'is homonymous. Those who have recently marked undergraduate examinations will know that, when the name 'Strawson' appears in answers on Kant the candidates probably have Peter Strawson in mind, while if it appears in answers on Hume they are thinking of Galen Strawson: if the name appears in answers on causality, or naturalism, it often seems likely that they do not know, or even care, that there are two prominent philosophers of that name.}

Cohen argues that the linguistic conception is preferable for several reasons. It allows us to prescind from the controversy over whether proper names have senses as well as references. It allows us to see that the capacity to re-use proper names is an immense convenience in natural languages. It prevents us from endorsing fallacious arguments that all true identity statements between names are necessarily true. It allows us to see that the actual way in which names function in natural languages is a form of indexical usage. It enables us to clarify the important distinction between what a speaker designates when uttering a name in a sentence and what the speaker has in mind or intends to designate, and it does so as part of a semantic account of names, not a pragmatic or speech-act one. The idiosyncratic conception does not achieve these important results: at best, it is an oversimplification of the function of names, which may have limited utility in some contexts but is quite wrong as part of a representation of natural language semantics.

The second half of this collection focuses on issues to do with knowledge, though there are obvious links to, and parallels with, papers 1-10. Papers 11-13 deal with some general epistemological issues, while papers 14-20 are principally concerned with issues of judgement under uncertainty – of induction, probability, confirmation and evidential support.

Paper 11 (*Third World Epistemology* [1985]) evaluates Karl Popper's hypothesis that there exists a third world, distinct from the material and the mental, of objective knowledge. Cohen points out that Popper is not here deploying some extended metaphor. His thesis is an ontological one: he claims to be a realist about this third world. Although the world is man-made, Popper insists that it exists to a large extent autonomously because it contains unintended by-products of our critical and creative thinking to date. These by-products and the existing products of our creative and intellectual activities are multifarious and diverse. Their realm can be said to be objective in that, when someone first formulates a problem which is a by-product of existing theory this process is, in effect, the discovery of a problem in the realm, not its creation.

Cohen draws attention to a number of major problems for Popper's hypothesis. First, since there are, or may be, barriers of time and space across which intelligent creatures cannot communicate, it will be impossible to know the extent of objective knowledge. Secondly, it would seem that the third world contains internally inconsistent theories, from which, according to classical logic, any arbitrary proposition (and its negation) can be derived. It is thus in danger of being uselessly overfull. Thirdly, Popper's conception sits uneasily with the way in which human knowledge normally seems to grow; and it is extremely unclear in what sense this growth, given Popper's hypothesis, is, as he calls it,

'Darwinian'. Cohen concludes that the hypothesis, once subjected to this critical analysis, contains far too many paradoxes and incongruities to be acceptable as an account of the development of our knowledge.

Paper 12 (*Guessing* [1974]) begins by noting that in recent epistemology and philosophy of science the concept of guesswork has been much used but seldom analysed, instead being treated rather vaguely as an active process (in contrast to passive doubt) which is more or less the same as conjecture. Cohen's aim in this paper is to provide an analysis of guessing by distinguishing it from conjecturing, speculating, suspecting, assuming, supposing, jumping to a conclusion and hypothesising. Such an analysis need not produce a grand overarching epistemological theory for it to have useful consequences in clarifying certain issues in epistemology.

Initially, we might suppose that 'guess' and 'conjecture' are simply a cognate pair of Latin-born and Anglo-Saxon-born terms for the same act. Such pairs are very common in English. {Though Cohen does not mention this, a seemingly redundant doubling of such terms is a familiar literary device – 'hire and salary', 'tomb enough and continent', for example –, in its subtler form being the figure of speech called hendiadys.} So, one might claim that there is no difference in meaning, merely in snob-value, between 'guess' and 'conjecture'. However, 'guess' seems sometimes to behave as a 'success-verb', sometimes with a neutral sense (in the way in which verbs like 'see' are often claimed to function) while 'conjecture' is never a success-verb. The terms also differ in their implications as to the adequacy or existence of the evidential grounds underlying them.

The paper works through a detailed comparison of the cluster of activities or states of mind listed above. Among the many points to emerge, three seem of immediate interest for contemporary epistemological debate. First, when one distinguishes guessing from suspecting it becomes clear that 'John is merely guessing that p' is compatible with, but neither entails nor is entailed by, 'John believes that p'. This undermines commonly made claims that knowing need not entail believing based on examples such as Radford's [1969] case of the nervous but successful examinee who thinks he is merely guessing. Secondly, we will be inclined to distrust a highly successful tipster who can point to no reasons at all for his even suspecting which horse will win, but claims instead to be endowed with great luck in guessing: this has consequences for any analysis of knowledge or rational belief in terms of the reliability of a source of information. Thirdly, noting that assuming differs from guessing in being more under one's voluntary control points up a distinction between a voluntary policy and an involuntary disposition, which seems

to be to some extent a precursor of – or is at least suggestive of – the distinction between acceptance and belief which is central to paper 20 and developed in Cohen's [1992] book *An Essay on Belief and Acceptance.*

Paper 13 (*Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses* [1979]) begins by noting that medicine, and in particular diagnosis, is an excellent live example of a field in which it is often inevitable that decisions should be made or inferences drawn under conditions of uncertainty. Two models exist for evaluating which diagnoses are preferable to others: one is explicitly and self-avowedly Bayesian; the other, Cohen argues, though not normally conceiving of itself as such, is 'Baconian' in the sense in which he deploys this term when analysing inductive probability or evidential support. The aim of this paper is to show that, although these models appear to be in competition, "when restricted to their legitimate roles they do not conflict with one another" (p. 175).

Central to the Bayesian model, in whatever precise form it takes, is the claim that diagnoses should be evaluated in terms of conditional probability (of the standard kind, elsewhere termed by Cohen 'Pascalian'). The goodness, or merit, or reliability of a diagnosis is identified with the probability that the patient has a particular disease given certain specified evidence. This probability can, at least in most cases, be calculated using the standard probability calculus. Cohen considers a number of objections to this model and offers answers to all but one: namely, that it ignores, and must ignore if it is committed univocally to the standard axioms of probability, the importance of weight of evidence. Generally, diagnoses cannot rationally be evaluated merely in terms of level of conditional probability without also taking account of the extent of the evidence on which the probability is said to be conditional.

The alternative model represents diagnostic reasoning either by decision-tables or flow-charts. Diagnosis selection then depends at least as much on the extent to which questions tabulated in the complete decision table have been answered as on mathematical probability. The evaluation of a diagnosis is more a matter of ranking its progress through a standardized sequence of questions – which depends upon gathering evidence – than of probability. This, Cohen points out, is just what also characterizes weight-based Baconian views of inductive support.

He concludes that the Bayesian evaluation measures how far evidence supports a diagnosis if it is assumed that this is all the available evidence: while a decision-table evaluation is Baconian in showing whether or not the balance of evidence obtained supports a given mathematical probability and, if it does, how much of the available evidence has been considered. Both kinds of judgement have their place in diagnosis.

In Paper 14 (*Are People Programmed to Commit Fallacies? Further thoughts about the Interpretations of Experimental Data on Probability Judgment* [1982]) the further thoughts are further to Cohen's [1979] paper but especially to his target article in *Behavioral and Brain Sciences* [1981]. The paper begins by distinguishing what Cohen calls the Preconceived Norm Method from the Norm Extraction Method. The former takes for granted a standard conception of probability in the light of which most people's probability judgements are erroneous: the latter starts from the assumption that, if one factors out performance errors, one can uncover an underlying competence in people's judgements provided these are taken to be judgements which need not correspond to the standard probability calculus. Cohen argues that the latter is clearly preferable. In its light, various alleged 'fallacies' – systematically irrational judgements by the subjects of experimental investigations into probability assessment – turn out to be explicable as rational, if not always accurate, applications of more than one conception of probability.

This paper considers four of these alleged fallacies (others are discussed in the earlier articles).

The first alleges irrational overconservatism in subjects' revisions of their prior probability estimates in the light of new evidence. However, the experimenters here are assuming both that one can assign prior probabilities directly in terms of given ratios and that posterior probabilities should be arrived at on the basis that the set-up is a chance one. But the first assumption ignores the subject's entitlement to take into account the reliability of the witness telling him or her the relevant ratios; and the second assumption ignores the reasonableness for the subject of being rather cautious as to assuming the situation to be one of pure chance – caution which becomes more reasonable as the task becomes more important.

The other three cases – the 'Gambler's Fallacy', judgements of the probability of a conjunction and awareness of the significance of the ratio of a sample size to population size – can be accounted for along similar lines. So, rather than Bayes's theorem not being applied, we can interpret the subjects' behaviour as consistent application of the theorem but with quite rational assumptions, different from those of the experimenters, about the prior and conditional probabilities appropriate to the set-up.

In general, Cohen concludes that left to their own devices subjects assume that causes are in play and then tend to construe the problem before them in Baconian terms. However, if they are prompted to approach the problem by being asked to assign a degree of belief or a betting-quotient or ratio, they attempt – in an untutored way – to make

judgements of Pascalian probability which, without suitable training, they often get wrong. But they should not thereby be taken to be using some intrinsically fallacious heuristic.

Paper 15 (*Inductive Logic* 1945-1977 [1980]) offers, after a brief historical preamble, a survey of the most influential mid-twentieth century accounts of inductive reasoning, followed by a summary of Cohen's own 'Baconianism'. The preamble very briefly presents a standard account of two approaches which emerged in the seventeenth century to the question of how one proposition may non-deductively support another. {I myself tend to think of these, perhaps rather crudely, as British and continental-European traditions respectively.} The former stems from Bacon's pioneering ideas about variative induction, when putative laws of nature become more and more securely established according as they incorporate an increasing variety of known uniformities. The latter is the mathematical calculus of chance, which had its beginnings in the work of Pascal and Fermat. In succeeding centuries Hume (to some extent), Whewell, Herschel and Mill developed Bacon's ideas, while mathematicians such as the Bernoullis and Laplace developed the theory of Pascalian probability (though, perhaps oddly, not until the twentieth century did it become a fully axiomatised calculus). These traditions developed pretty well independently until the mid-nineteenth century: only then did the assumption emerge that inductive support ought to conform to Pascalian principles. From the 1920s onwards this assumption was largely unquestioned, even though its consequences produce difficulties for the most influential attempts at developing inductive logic.

After first outlining the three most famous paradoxes or puzzles to have influenced theories of induction – Hempel's paradox of the ravens, Goodman's 'grue' problem and Kyburg's lottery paradox – Cohen expounds and criticises in turn the positions developed by Carnap, Hintikka, Levi and Hesse. He then asks how inductive logic might instead be developed if, in the spirit of Baconian tradition, it avoided making the 'Pascalian' assumption described above. The inductive reliability of a hypothesis, he claims, can be ranked (though, importantly, not measured) by reference to its performance on a series of increasingly stringent tasks. Given this 'method of relevant variables', inductive support can be seen not to be identical to, nor capable of being mapped onto, Pascalian probability. This is so because, it can be shown, Baconian inductive support can satisfy neither the multiplication nor complementarity axioms of Pascalism probability.

Paper 16 (*Some Historical Remarks in the Baconian Conception of Probability* [1980]) discusses at greater length the tradition running from Bacon to Mill which the previous paper presented in a very compressed

way. Its conclusion is a natural extension of the final point above. Although inductive support is not conformable to Pascalian probability, Cohen claims that it may legitimately be termed inductive probability, though non-Pascalian (the analogy is with non-Euclidean geometry). For, he argues, there is a strong case for a Baconian interpretation of probability as understood within Anglo-American law when it is used to state standards of proof (proof in the juridical sense, that is). There is a similarly strong case when we consider historical research. Most importantly, he argues – echoing his views on meaning expressed in the first half of this volume – the use of the term 'probability' in these disciplines cannot be put down to accident or mere homonymy. Instead, if – as he argues elsewhere and many other philosophers have held – probability is a generalization of provability, the resulting semantics for the term 'probable' generates both Pascalian and Baconian readings for that term. His pluralism does not, therefore, avoid or give up on a unifying justification for the systematic ambiguity of 'probability'.

Such is the conclusion of the paper. En route towards it, he defends Bacon's own views against misinterpretations, presents the views of his seventeenth-century admirers (who included Hooke and Boyle) and discusses Bacon's influence on Hume, Mill and Herschel. The discussion of Hume concentrates on a distinction, central to his account of probability, which has failed to attract the attention it merits (as Cohen points out, even Hacking's now-classic text *The Emergence of Probability* fails to mention it). This is the distinction between probabilities arising from analogy and probabilities arising from chance or cause. Hume adopted a classical equipossibility approach to the evaluation of probabilities arising from chance and a "straight rule" approach to those arising from cause, and saw little difference between them – both could be measured numerically and obeyed Pascalian laws. But he also thought there was a quite different type of probability, analogical probability. Famously, of course, he held that any reasoning about causes and effects relied on two factors: a constant union of one type of situation with another type, and the resemblance of a present situation to one of them. If either factor is diminished then our feelings of necessity are reduced to probability. Less than wholly constant union generates probabilities arising from chance or cause: less than perfect resemblance generates analogical probability. He regarded the latter as admitting of degree, but he never suggested, as he did for the former, that it could be measured. So, although Hume himself did not take the final step to arguing that they could not be measured, it is natural to understand his analogical probabilities as Baconian.

Paper 17 (*Twelve Questions about Keynes's Concept of Weight* [1985]) provides a detailed analysis of a concept which Keynes introduced in developing his logical-relation theory of probability but did not himself – since he could see no practical significance to it – analyse further. For Keynes, the probability of hypothesis H on evidence E is a property of the argument from E to H. It measures the balance between the favourableness and unfavourableness of E to H. As the relevant evidence at our disposal increases, this probability may increase or decrease: but, he says, something else seems to have increased in either case – the basis on which we rest our conclusion. Thus, there is a parameter other than probability which, roughly speaking, measures the sum rather than the balance of the favourable and unfavourable evidence: he calls this the 'weight' of the argument. Cohen's aim in this paper is to show why, *pace* Keynes, a theory of weight is needed and to ask what form it should take if it is to preserve Keynes's original intuition. First, he notes that – unless we adhere to a logical-relation theory – we have an obvious concern (both practical and epistemic) to ensure that our probability judgements take account of as large a sum of relevant evidence as possible, given a range of constraints on the exhaustiveness of enquiry. So we do value weight of evidence. How, then, is it to be understood? Cohen probes the issue of how weight is increased by new evidence: we require rather a subtle account of relevance in order to answer this question. He asks how arguments can be compared in respect of their weight. He points out that, although other philosophers have used the term 'weight' to denote some form of probability-related measure, this is obviously not the sense Keynes had in mind: for him, arguments may have high probability but low weight, or vice versa. It is true, though, that although we can know the weight of an argument without knowing its probability, we cannot make any use of our knowledge of the weight of an argument unless we have some knowledge of its probability.

Finally, Cohen points out that although the analysis of weight which he is offering is quite independent of the account of Baconian legisimilitude which he develops elsewhere [1970, 1977] "both lines of reasoning converge on precisely the same underlying structure" (p. 277). He concludes that Keynesian intuitions about probability and weight may well lead to the same destination as Baconian intuitions about causality.

Paper 18 (*Some Steps Towards a General Theory of Relevance* [1994]) begins by criticizing Keynes, Carnap and their followers for attempting to reduce the concept of relevance to that of evidential relevance, explicable in probabilistic terms. He insists that their analysis does not extend to a satisfactory analysis of legal, moral or conversational relevance. As in other contexts, he rejects the suggestion that there is no conceptual

unity underlying these uses of the term, rather just a set of accidentally associated homonyms. His aim in this paper is to characterize, via close attention to examples, the unitary underlying structure of the concept of relevance: the existence of such a structure is perfectly compatible with their being a variety of criteria for applying the concept.

He begins by noting that many sentences about relevance are ambiguous between conversational and non-conversational relevance. Once disambiguated, it is evident that no analysis in terms of differences between prior and posterior probability could be appropriate to the conversational reading. What conversational and non-conversational relevance have in common is that both relate the truth of a proposition (in the latter case, but not the former, one stating the occurrence of a sentence) to some corresponding question, that is, issue or problem. The general structure underlying relevance derives from the nature of this relation: it is, in all cases, one of the proposition's providing some reason for accepting or rejecting an answer to that question. This may be an evidentially based reason (which could include causes) or a legal or moral reason, or a conversational ground. More precisely, Cohen concludes:

> "...a true proposition R is non-conversationally relevant to an askable question Q if and only if there is a proposition A such that the truth of R is or would be some reason, though not necessarily a complete or conclusive reason, for anyone's accepting or rejecting A as an answer to Q, where this reason is warranted by some non-conversational principle". [p. 279]

For evidential relevance this principle may be expressible probabilistically, or it may derive from a set of covering laws. But, Cohen concludes, this type of relevance is in no way more central than relevance where the warranting principle is a moral principle, a legal rule and so on.

Paper 19 (*Should a Jury Say What it Believes or What it Accepts?* [1991]) applies a distinction between belief and acceptance, alluded to earlier in this introduction, which Cohen developed in his [1989] and [1992], to a normative issue about judicial decisionmaking. For Cohen, believing some proposition is an involuntary disposition to feel it to be true (of course, this feeling may come in varying strengths). Accepting a proposition, in contrast, is a voluntary policy of taking the proposition as a basis for inference or action. Much of what we believe we also accept, of course, and vice versa. Belief is often a reason for acceptance; acceptance often promotes or causes belief. But there are important contexts in which they do not coincide: for instance, it is part of an advocate's duty, in an adversarial system, to accept his client's story and put it to a tribunal as powerfully as possible even if he may not believe a word of it (assuming his client has not admitted it to be a fiction!). Cohen next discusses how different legal systems actually do, or seem to, require their triers of fact to arrive at verdicts, before asking the title question: should – in terms of justice and the public good –

belief that p be necessary and/or sufficient for a verdict that p? He points out that the requirement of swearing an oath implies that what ought to be crucial is acceptance, not belief: one cannot commit oneself to do something over which one has no voluntary control. Further, the existence of extensive means of regulating a jury's response to evidence – for example, via exclusionary rules – does not fit well with the view that verdicts should express belief rather than acceptance. If we may take it that the practice of oath-taking and the exclusion of prejudicial evidence are grounded in promoting fairness of verdicts, we may take it that both exist to prevent passive surrender to belief and do so on the grounds that this would be wrong – unjust or injurious to the public good.

Finally, Cohen turns to the objection that since legal standards of proof come in different strengths, sometimes expressed as probability, and acceptance is an all-or-nothing matter while belief is gradable, it follows that verdicts either do not and should not be based on acceptances, but on beliefs. Cohen's response is to suggest that if judicial probability, as it applies to ultimate issues, is to be interpreted as Pascalian probability (which he himself does not accept) it can only be so under a subjectivist interpretation which, examined closely, turns out to be better given an acceptance-based rationale rather than its conventional expression in terms of partial beliefs. So, he argues, this objection has no force.

Paper 20 (*Are there Ethical Reasons for Being, or not Being, a Scientific Realist?* [1997]) addresses, in a different context, a topic implicit in much of the previous paper. As of the law, so too of science various kinds of normative question may be asked. One kind may be termed 'epistemological': for instance, how should theorizing proceed, or experiments be designed, so as to lead to knowledge? Another kind may be termed 'ethical': for instance, what good can science achieve?; ought research to be constrained by consideration of the public good? In this paper Cohen focuses on the issue of the relationship between these two kinds of question, specifically with regard to whether there can be ethical reasons for adopting a realist view of science.

He begins by assessing an argument that there can be no such reasons, which relies on two premises: that all ethical judgements in science involve the application of a more general moral principle to the particular circumstances of science, and that scientific realism lacks the logical structure needed for the application of such principles. He accepts both premises, and that they appear to entail a negative answer to the title question. However, he points out, the argument works only on the further assumption that science is not independently valuable: and, if

we conclude that it is, a realist position seems altogether more appropriate than an instrumentalist one, since the latter implies that science should be pursued for purposes other than the attainment of knowledge for its own sake. If we believe that scientific theory-building is an activity which should be pursued because it has intrinsic value, it seems distinctly odd to suppose that those theories are merely predictive tools. And, he claims, there is quite a strong case for the claim that science, like art, does have intrinsic value. Finally, he argues that, if the conflict between realism and instrumentalism is seen as a dispute about ontology, the positions collapse into one another. Instead, it must be seen as an epistemological clash. There can then be seen to be methodological reasons for being a realist. If one adds the plausible claim that there is a general moral duty to follow the best methodological principles in one's work to this claim that realism is methodologically preferable to instrumentalism, it follows that there is a professional duty (a derivative moral duty) to be a scientific realist.

1

ON THE PROJECT OF A UNIVERSAL CHARACTER

I

During the last thirty years or so the practice has grown up among logicians of attributing the project of a universal character to Leibniz alone among seventeenth century thinkers. This attribution is to be found, for instance, in L. S. Stebbing's *Modern Introduction to Logic*,[1] in Cohen and Nagel's *Introduction to Logic and Scientific Method*,[2] in M. Black's *Nature of Mathematics*,[3] in J. H. Woodger's *Axiomatic Method in Biology*,[4] and in 0. Neurath's introductory article in the *International Encyclopaedia of Unified Science*.[5] And it dates, I suspect, from the publication of C. I. Lewis's *Survey of Symbolic Logic* in 1918. Lewis mentioned that Leibniz acknowledged a debt in this connexion to Raymond Lully, Athanasius Kircher, George Dalgarno and John Wilkins. But he considered their writings contained "little which is directly to the point".[6] In this Lewis was obviously right with regard to Leibniz's conception of a calculus of reasoning, but wrong, as I shall try to show, with regard to the project of a universal character, which seems in fact to have been an intellectual commonplace in seventeenth century Western Europe. This somewhat neglected by-way of philosophical history is worth a brief review, I think, not only in order to fix more precisely the respect in which Leibniz was the only seventeenth century precursor of modern symbolic logicians, but also because it draws attention to an early widespread philosophical muddle about the construction of artificial languages.

Lewis's summary of Leibniz's own ideas, based on L. Couturat's *La Logique de Leibniz*, is quite sufficient to show how much Leibniz's project

had in common with those of previous thinkers on the subject. His universal character, in which he was enthusiastically interested from the age of eighteen (1664), was to fill three main roles. It would be what modern linguists call an "international auxiliary language", enabling men of different nations to communicate with one another. It would provide what Lewis calls a "logistic" treatment of science in general, a simplified system of symbolism for the exact expression of all actual and possible knowledge. And it would serve as an instrument of discovery and demonstration. The first two purposes could be achieved on Leibniz's view by devising a notation in which each single or basic symbol represented a simple concept and complex notions were expressed by combining the appropriate symbols in one way or another. The number of fundamental symbols would thus be as small as the number of primitive concepts in human thought, so that the notation, together with its vocalisation, could be learnt in a few weeks. Moreover, throughout the world science would be developed in the form of a single, unified encyclopaedia, scientists would be able to attain the same degree of rigour in metaphysics and morals as in geometry and analytics, and nothing chimerical would ever be written down. However, Leibniz never did much himself to develop this system of symbolism though he was constantly urging others to do so. He elaborated only a few of the many analyses of complex notions which would be required in order to establish which concepts should be regarded as primitive. And he learnt enough about two systems of notation to see that they would not serve his purpose. The earlier of these was based on the division of concepts into classes in accordance with their degree of complexity. Any concept could then be symbolised by a fraction of which the denominator indicated the number of the class and the numerator the number of the concept in that class. Later Leibniz considered symbolising primitive concepts by prime numbers and complex concepts by the appropriate product, so that logical synthesis would be represented by arithmetical multiplication and logical analysis by resolution into prime factors.

But besides being an international and scientific language Leibniz's universal character was also to serve as an instrument of discovery and demonstration by exhibiting the implications of what was already known. It was to see into the "inner nature" of things like a "new telescope", and guide our reasonings like an "Ariadne's thread". Leibniz attempted to devise a calculus which would operate on the formulae of his character to this end. It is these fragments of a calculus, expressed in a notation of concept variables and logical constants, and not at all restricted in possible application to Leibniz's own character, which justify the claim that Leibniz anticipated Boole. I shall try to show that, apart from some mi-

nor details, it is only this conception of a logical calculus as an ancillary to his universal character which distinguishes Leibniz's thought on the subject from that of some of his contemporaries and predecessors. Several of these projected, and some completed, a universal character which they hoped would fill the same three roles as Leibniz had in mind, constituting at once a medium of international communication, a simplified notation for science, and a method of discovery and demonstration.

II

The need for an international auxiliary language was widely felt in seventeenth century Europe for several reasons. The learned were increasingly using their own several vernaculars instead of Latin, merchants faced an immense linguistic problem on the new trade-routes, and missionaries were confronted with the same difficulties. Consequently the recent discovery that written Chinese was used in the Far East as a means of intercommunication by peoples whose spoken languages differed greatly from one another was of great interest. Indeed, it set the pattern for almost all seventeenth-century speculation on these matters, and was referred to by almost all who discussed them. The ideal was not a new spoken language which could also be written (like Lingua Franca or Pidgin-English) by a system of phonograms, but a written language constituted by a new system of ideograms which could also be spoken. As such it was intended to be something quite different from the ciphering of words or letters practised by cryptographers even in the ancient world. The programme for its construction was first enunciated by Francis Bacon in *The Advancement of Learning*[7] in 1605, where he speaks of the project as "the mint of knowledge" since "words are the tokens current and accepted for conceits, as moneys are for values". What he seeks is a system of "real characters" which, as in China, would "express neither letters nor words in gross, but things or notions" and would be "as many, I suppose, as radical words".

Bacon mentioned this project again in *De Augmentis Scientiarum*[8] in 1623, where he added that "any book written in characters of this kind can be read off by each nation in their own language". But he was not enthusiastic over its utility, apparently because of the inconveniently large number of characters he thought it would require. Descartes thought he could circumvent this difficulty when Mersenne informed him in 1629 of a similar project, formulated apparently by a Mr. Hardy.[9] As soon as "the true philosophy" was known, the "clear and simple ideas" which are the basis of good science could be discovered, enumerated and arranged. It would then be possible to devise a system for symbolising these so that the most complex notions could be as easily expressed

as the largest numbers, because an order would have been established among all possible human thoughts similar to that prevailing naturally among members, and the whole language could be learnt in a few days. Above all, this language would represent everything so clearly to the human judgement that error would be almost impossible and even peasants would be better able to assess the truth than can philosophers at present.

Leibniz actually took a copy of this passage,[10] and added to it the comment that "although this language depends on the true philosophy it does not depend on its perfection". It could be established while philosophy was yet imperfect and grow as knowledge grew. "Meanwhile it would be a wonderful help for preserving what we know, seeing what we lack, discovering the means of attaining this, and above all for settling controversies where the correctness of a chain of reasoning is at stake". But there were others before Leibniz who thought that the construction of a universal character on mathematicist principles need not wait for the scientific millennium. Indeed the project of a universal character of some sort became a commonplace from Descartes' time onwards, and the languages seem to have been designed as often on Baconian as on Cartesian lines. To the former type probably belonged the scheme drawn up by William Bedell, Bishop of Kilmore, in 1633, which he persuaded a Reverend Johnston to execute. The results of their work were unfortunately lost in the Irish rebellion of 1641.[11] Of this type too were the projects of the elder Vos and Herman Hugo,[12] and also those of Philip Labbe[13] and Edward Somerset, second Marquis of Worcester.[14] At Ipswich in 1657 Cave Beck published a grammar and dictionary of such a language, entitled *The Universal Character, By which all the nations in the world may understand one another's Conceptions, reading out of one Common writing their own Mother Tongues.* In his preface he remarked "This last century of years, much hath been the discourse and expectation of learned men, concerning the finding out of a universal character, which if happily contrived, so as to avoid all equivocal words, anomalous variations and superfluous synonomas ... would much advantage mankind in their civil commerce, and be a singular means of propagating all sorts of learning and true religion in the world: such a character being to be learned in as few weeks as the Latin tongue usually requireth years to be perfect in." Similar works were published by J. J. Becher at Frankfurt in 1661 (*Character, pro notitia linguarum universali*), and by Athanasius Kircher at Rome in 1663 (*Polygraphia Nova et Universalis*), who provided dictionaries into and out of his language for Latin, Italian, French, Spanish and German. Beck, Becher and Kircher all make much use of numerals in their notation, which is mostly built up by num-

bering English or Latin words in their alphabetical order; and Becher and Kircher acknowledge a debt to the cryptographic researches of Johann Tritheim, published early in the sixteenth century under the title of *Polygraphia.* The syntax, accidence and word-order of their languages is, roughly, the highest common factor of Romance, Teutonic and Semitic grammars, and their dictionaries presuppose a one-one correspondence of meaning between the vocabularies of all ordinary languages. So that in effect their languages are systems for ciphering a limited group of languages on a unitary pattern. No Japanese, for instance, could read his vernacular directly out of their sentences, though he could no doubt learn to translate them. But each of these authors nevertheless claimed to be expounding a universal real character, in Bacon's sense (like the modern international code of nautical signals) and not a mere cipher of limited application.

Mersenne seems to have been the first to design a universal language of the Cartesian type, for he mentions this achievement in a letter of 1636 or 1637.[15] He was soon followed by F. Lodwick,[16] a London merchant, J. A. Comenius,[17] the Czech educationalist, and Sir Thomas Urquhart, the translator of Rabelais. Mersenne did not describe his system in detail, and there seems no evidence that Comenius ever worked out any details. But Lodwick wrote two books on the subject. The first of these was published in 1647 as *A Common Writing, whereby two, although not understanding one the other's language yet by the help thereof may communicate their minds one to another* and the second, which was largely a refinement of the first, appeared in 1652, as *The Groundwork, or foundation laid (or so intended) for the framing of a new perfect language and an universal or common writing.* Lodwick's characters are made up of signs for a large number of basic roots and of "distinctional marks" around each sign which signify the derivative form required. Some of these derivative forms, called "abbreviatives", consist of a small range of variations around each radical notion. Its contradictory, for instance, can be signified in this way. The other derivative forms arise from the inflexions necessary to produce a verb, noun, adjective and adverb, with a large range of moods, tenses, cases, etc., from each abbreviative. The resulting scheme is tidier and more systematic than, say, Beck's. And in 1653 Lodwick was clearly of the opinion that "the proper names of things to give them signification is the work, we suppose, of a sound philosopher, who from the knowledge of things and their order in nature, should give them names accordingly, describing that in them by their name, by which in the naming they may be known". Lodwick was convinced too that his character would "much assist the true knowledge of things which is at present much hindered" by verbal ambiguity and

vagueness, and he pointed out how it would be of greater use to science than to poetry or rhetoric.[18] But his published scheme was still very far from the degree of simplification which Descartes thought possible. Moreover, it was only an outline, for Lodwick did not, like Beck, provide a lexicon for his Language.

Although Lodwick was interested enough in the problem of phonograms to devise a new phonetic notation, which was published in the *Philosophical Transactions* of the Royal Society in 1686 under the title of "An Essay towards an Universal Alphabet", his universal language was ideographic. And Urquhart seems to have been the only seventeenth-century thinker who based his new language on phonograms. Unfortunately most of Urquhart's manuscript was lost at the battle of Worcester in 1651. But an "Introduction to the Universal Language" was saved, and published the next year. [19] Urquhart claimed that he had "couched an alphabet materiative of all the words the mouth of man, with its whole implements, is able to pronounce", and brought "all these words within the system of a language, which, by reason of its logopandocie, may deservedly be intituled The Universal Tongue". So that "there is not a word utterable by the mouth of man, which, in this language, hath not a peculiar signification by itself". Indeed, "as, according to Aristotle, there can be no more worlds but one, because all the matter whereof worlds can be composed is in this; so there can be no universal language but this I am about to divulge unto the world, because all the words enunciable are in it contained". Yet "this world of words hath but two hundred and fifty prime radices upon which all the rest are branched". Urquhart had noticed that "no language ever hitherto framed hath observed any order relating to the thing signified ...; for if the words be ranked in their alphabetical series, the things represented by them will fall to be in severall predicaments; and if the things themselves be categorically classed, the words whereby they are made known will not be tyed to any alphabetical rule". This "imperfection", however, would be remedied in his own language, where there would be "a proportion between the sign and the thing signified". Accordingly, among the sixty-six advantages which Urquhart claimed for his scheme, he asserted that "sooner shall one reach the understanding of things to be signified by the words of this language, than by those of any other, for that as logarithms in comparison of absolute numbers, so do the words thereof in their initials respectively vary according to the nature of the things which they signify". Moreover, "for definitions, divisions and distinctions, no language is so apt; ...for the affirmation, negation and infinitation of propositions, it hath properties unknown to any other language, most necessary for knowledge ; ...and ...in matters of

enthymems, syllogisms, and all manner of illative ratiocination, it is the most compendious in the world".

In 1653 an anonymous Spaniard is said to have published at Rome a scheme for a universal language on Cartesian lines, which he proposed to achieve by dividing all concepts into classes and sub-classes and numbering both the concept-class and the concept within its class.[20] And by 1654 a universal character on Cartesian lines had also been projected by Seth Ward, professor of astronomy at Oxford and previously a lecturer in mathematics at Cambridge, who thought it "a misfortune to the world, that my Lord Bacon was not skilled in mathematics, which made him jealous of their assistance in natural enquiries". Ward, like Descartes, thought that the number of different characters requisite for a universal language on Baconian lines would be so large "that the tradition of learning or facilation of it would be but little advanced by this means. But it did presently occur to me, that by the help of logic and mathematics this might soon receive a mighty advantage, for all discourses being resolved in sentences, those into words, words signifying either simple notions or being resolvable into simple notions, it is manifest that if all the sorts of simple notions be found out, and have symboles assigned to them, those will be extremely few in respect of the other, ... the reason of their composition easily known, and the most compounded ones at once will be comprehended, and yet will represent to the very eye all the elements of their composition, and so deliver the natures of things: and exact discourses may be made demonstratively without any other pains than is used in the operations of specious analytics."[21]

Ward never developed these ideas in detail, but two friends of his did–George Dalgarno and John Wilkins–though the extent of their debt to him and to each other is in doubt.[22] Dalgarno, an Aberdonian who taught for thirty years at a private grammar-school in Oxford, had been maturing his thoughts on the subject during the later fifties. In a letter to Samuel Hartlib (the publisher-philanthropist and friend of Comenius), for instance, dated 20 April 1657, he criticises–quite rightly–Cave Beck's Universal Character as "nothing else but an enigmatical way of writing the English Language", and speaks more favourably of a "treatise" published in 1647, which is probably Lodwick's *Common Writing*.[23] He published his own *Ars Signorum, vulgo Character Universalis et Lingua Philosophica*, in 1661. In this work he drew up a "Lexicon Grammatico-Philosophicum, seu tabulae rerum, et notionum omnium simpliciorum, et generaliorum, tam artefactarum quam naturalium, rationes et respectus communiores, methodo praedicamentali ordinatas, complectentes". There were seventeen irreducible categories according to Dalgarno, and he designated them by seventeen letters of which each was the initial

letter of all names falling under one category. The second and third letters of each name signified the sub-class and sub-sub-class to which it referred, and there were provisions for further specialisation where necessary, for instance, in order to deal with the classification of plants. Certain letters could also be prefixed to a word in order to denote its opposite, the mean between two extremes, etc. The language drew no fundamental distinction between different parts of speech, for its author held that all these, even prepositions and interjections, could be derived from his primitive notions (though he made an exception here in favour of six pronouns). Instead, there were rules of word-order designed to show when and how a word was being used adjectivally or adverbially; and inflexions of tense, mood and voice were signified by special suffixes. Dalgarno claimed to have discovered by a process of "analysis logica", applied to ordinary English and Latin usage, that his system adequately represented the normal articulation of human thought. In this respect it is a crude anticipation of the analysis of conversational language practised by symbolic logicians like Reichenbach. But Dalgarno's system is also a precursor of what Lewis calls "logistic". Though he criticised some features of Descartes' philosophy he was thoroughly imbued with its mathematicist spirit. He attacked the Baconian theory, popular among many of his experimentalist contemporaries, that there was no use for a "doctrina praedicamentalis", an arrangement of all knowledge in an ordered system. He held that, while metaphysics and logic constituted a single art, the difference between this art and linguistics was precisely analogous to the difference between the symbolised and its symbol, so that when the two were correlated there would be but one science of them all. Accordingly, Dalgarno believed not only that his universal character would make it much easier for the young to learn the principles of true science and the correct practice of logic, but also that the analyses which it made possible would lead "in penitiorem et interiorem rerum cogrutionem". And in view of the widespread contemporary interest in stenography and in the physiology of speech, sight and hearing, it is hardly surprising that Dalgarno made two further claims for his Character.[24] He asserted that it produced the best short-hand so far invented, and also that it could be used for conveying knowledge to the deaf and dumb. (His specialised work on the latter subject, *Didascalocophus, or the Deaf and Dumb Man's Tutor*, was not published till 1680.)

John Wilkins, Cromwell's brother-in-law and first co-secretary (with Henry Oldenburgh) of the Royal Society, had published as long ago as 1641 a book on cryptography in which he devoted one chapter to the problem of a universal character.[25] In 1668 he published his *Es-*

say towards a Real Character and a Philosophical Language. This work had been commissioned some time previously by the Royal Society, and would have been published much sooner if London's great fire of 1666 had not destroyed most of the printed sheets and a great part of the manuscript. As befitted its sponsors it was a much larger and more thorough undertaking than Dalgarno's, though on the same general lines. Wilkins was aided in it by three collaborators, expert in botany, zoology and lexicography, respectively. There were now forty irreducible categories or *summa genera* and a vast range of subordinate classifications. A much wider range of ordinary linguistic usage was examined, and as a result Wilkins adopted a distinction, which Dalgarno, though he had himself once drawn it, had come to the conclusion was not fundamental. Wilkins firmly distinguishes in his language between what he calls "integrals" (corresponding to nouns, verbs, adjectives and adverbs) and what he calls "particles" (corresponding to pronouns, conjunctions, prepositions, inflexions, interjections and the copula). There was a much more comprehensive "alphabetical dictionary, wherein all English words according to their various significances are either referred to their places in the philosophical tables or explained by such words as are in those tables". Wilkins' character, like Dalgarno's, could be vocalised in the form of a distinct spoken language, as well as translated into any written or spoken vernacular. But Wilkins also devised an entirely new notation (looking on a printed page something like a cross between Pitman shorthand and ancient Syriac) for writing his character.

Wilkins believed in the same two fundamental requisites for the construction of a universal character as did Dalgarno. The first necessity was "so to contrive the enumeration of things and notions, as that they may be full and adequate, without any redundancy or deficiency as to the number of them, and regular as to their place and order". The second basic requisite was so to contrive a mark for every thing and notion that these marks should have "a dependence upon, and relation to, one another, as might be suitable to the nature of the things and notions which they represented", and that thus, "besides ... helping the memory ..., the understanding likewise would be highly improved, and we should, by learning the character and the names of things, be instructed likewise in their natures". Indeed, he felt convinced that his system would "contribute much to the clearing of some of our modern differences in religion, by unmasking many wild errors, that shelter themselves under the disguise of affected phrases, which being philosophically unfolded, and rendered according to the genuine and natural importance of words, will appear to be inconsistencies and contradictions. And several of these pretended, mysterious, profound notions, expressed in

great swelling words, whereby some men set up for reputation, being this way examined, will appear to be either nonsense or very flat and jejune". Thus, while Hobbes (as a result of his knowledge about Chinese) had achieved an *ad hoc* linguistic critique of ontology, for instance, by asking whether those who spoke a language which lacked a verb "to be" would ever use or be worried by the notions of "entity", "essence", and the like,[26] Wilkins was apparently proposing to achieve similar deflationary purposes by systematic analysis in terms of a philosophical language. Here again he was following Dalgarno, who had claimed for his own character, in a dedicatory letter, that "veritates philosophicae variae aperientur, v.g. Nullas dare notiones transcendentales, vel extravagantes, ut de uno, vero, bono, fabulantur, potius quam philosophantur metaphysici". And, like Dalgarno, Wilkins shared neither Descartes' belief that the construction of such a universal character must await the scientific millennium nor Urquhart's self-confidence that he was the linguistic messiah. He hoped that other members of the Royal Society would develop and improve his system. In order to examine its utility the Society appointed a special committee, among the members of which were Robert Boyle, Christopher Wren, John Wallis and Robert Hooke.[27]

III

In view of Leibniz's great interest in the details of Dalgarno's and Wilkins' universal characters it is surprising that he seems to have read their books rather carelessly in one respect. To an example of Dalgarno's character which belonged to him Leibniz appended the comment[28] that, as he "had said to Robert Boyle and Henry Oldenburgh", Dalgarno and Wilkins had not sufficiently grasped the magnitude of the thing and its true use, "for their language or character achieves this alone, convenient communication between those sundered by language, but the true real character, as I conceive it, would be thought one of the most apt instruments of the human mind, with an invincible power for discovery, memory and judgement". Leibniz's misunderstanding in this respect was echoed in 1901 by L.Couturat in his comprehensive study of Leibniz's logic which immediately became the standard work of reference on the subject.[29] It may be that Lewis's error arose from relying too much on Couturat's thoroughness. I hope, however, to have made it clear that in conceiving of a universal character as an important new intellectual tool, with powers not possessed by ordinary languages, Leibniz was anticipated not only by Wilkins and Dalgarno, but also by Ward, Urquhart and Descartes. It is only in thinking of a logical calculus as an ancillary to his universal character which would be useful for eliciting implications that Leibniz seems to have had no precursor. Yet Leibniz, as is well

known, was inspired in this direction, not only by reading Hobbes' elliptical remark "Per ratiocinationem intelligo computationem", but also by his early acquaintance with Raymond Lully's *Ars Magna* which Kircher had revised and developed by 1665.[30] Kircher's system, like Lully's, comprised a table of fundamental notions, specially symbolised, from which the permissible combinations of symbols could be read off as theorems. But apparently Kircher did not associate his own universal character or polygraphia with this scientific method. After the early publication of his *de Arte Combinatoria* (1666), in which he sought to develop the ideas of Lully and Kircher, Leibniz shows, in several discussions of his project for a universal character,[31] too much respect for the heuristic importance of experiment to expect a complete answer to all scientific enquiries along the lines of Lully's method. But in conformity with his doctrine that the predicate of any true proposition is "contained" in its subject, he was far from condemning Lully's art, as Bacon did, as "a method of imposture ..., being nothing but a mass and heap of the terms of all arts, to the end that they who are ready with the terms map be thought to understand the arts themselves".[32]

In conclusion it may be of interest to review briefly the main reasons why Bacon was correct in his low estimate of the prospects of a universal character, and why the seventeenth-century mathematicists' project of one which should constitute at once a medium of international communication, a simplified notation for science, and a method of discovery and demonstration, was a philosophical muddle and never of the slightest direct help to any merchant or scientist, even though it may indirectly have done something to inspire the later work of men like Linnaeus and Lavoisier. No doubt Kepler's laws, for instance, were not deducible within any of the proposed systems. No doubt a Cartesian misconception of the logic of scientific argument, and an Aristotelian preoccupation with the problems of classification rather than the study of motion, were often major factors in the muddle. But there were also other factors. It was clearly a mistake to think that the same language could serve adequately both as an unspecialised international auxiliary and also as a scientific terminology. J. R. Firth[33] has claimed, indeed, that Descartes envisaged two kinds of universal language, a philosophical one based on a logical system and an ordinary one for "esprits vulgaires". But Descartes' remarks on the subject do not lend themselves to this interpretation. He clearly thought, like Urquhart and the others, that any language would be far easier to learn if it were isomorphic with the facts of nature as systematised by science. Dalgarno even reproduced in one of his broadsheets a certificate, dated May 1658, from Richard Love, then professor of divinity at Cambridge, to the effect that two young

bachelors of arts, after spending two hours a day for a fortnight in the study of his character, had learnt to communicate with one another in that language and to translate anything into, or out of, it.

Now, the structural simplification which serves an international auxiliary best, of course, is a sweeping abolition of grammatical inflexions–a simplification which all these languages were far frcm carrying out. But, when words are constructed on a system parallel to the scientific classification of the things to which they refer, their meanings are more, not less, difficult for nonscientists to avoid confusing. Moreover, though naming, description, classification, generalisation, reasoning, explanation and prediction are the main tasks of scientific speech and writing, no-one would now deny that these are far from being the only purposes which we require an everyday language like an international auxiliary to serve. But the range and flexibility of meaning which we expect frcm an unspecialised vocabulary is quite inappropriate for a taxonomic notation, say, like that of modern chemistry. Again, scientific theory is best served by a language which carries over the least possible irrelevant associations from non-scientific discourse. But an international auxiliary is most easily learnt if its radicals are already familiar and these are therefore best drawn from words already in international currency, as in L. Hogben's Interglosa.

Finally, while an unspecialised language must be capable of application to all fields of human experience, it has yet to be shown that the ideal of a systematic terminology can usefully operate not only at the level of each individual science but also at the level of science as a whole. We can organise the richness of our non-scientific vocabulary into tables of synonyms, antonyms, etc., like those provided by Roget's *Thesaurus*, which acknowledges Wilkins's priority in its field, and this may be useful for writers or cross-word solvers, say, who are at a loss for a word. Or we can minimise the number of words and grammatical rules in an unspecialised language, as in Basic English, and this may make it easier for foreigners to learn. Or we can design language patterns in the course of what R. Carnap[34] calls "pure syntax" and "pure semantics", and from the study of these formal structures, as from other methods of analysis, we may be able to discern logical principles–"general syntax"–underlying all possible languages. Or we can construct an interpreted axiom-system for some part of natural science, as J. H. Woodger[35] has tried to do for embryology, genetics and biological taxonomy. This, as Woodger claims, may for a time give order to the experimentalists's records and direction to his activities. But if we try to do all these four things at once we may fail as badly as Dalgarno and the others failed, since projects like Roget's *Thesaurus* and Basic English, on the one hand, and Carnap's

or Woodger's constructions, on the other, have little in common. The former are bound to accept unquestioned the tasks which a language is called upon to perform and are concerned instead to regulate in one way or another the shapes and sounds with which it performs them, while the latter are interested primarily in systematising the tasks which some or all languages can or should perform and are comparatively indifferent to the shapes and sounds which are used for this. Moreover, unless the unity of science, in some important sense, can be shown to be more than a metaphysical speculation or a restrictive programme, it is not surprising to find that in practice scientists prefer to follow Bacon's principle that "uniformity of method is not compatible with multiformity of matter".

Accordingly, the impossibility of combining a practicable international auxiliary with an adequate scientific terminology has been stressed in this century both by linguists and by scientists.[36] When, for instance, Peano published the fifth volume of *Formulaire de Mathématiques* (1908) in Interlingua it was only his meta-language that was this uninflected Latin. But, since some philosophers[37] are still prone to talk about the properties of "a perfect language" as if the criteria of linguistic perfection were a uniquely determined set of compatible ideals, it is perhaps worth while pointing to the errors of the seventeenth century as an illustration of this phrase's misleading character.

Notes

1. pp. 483 ff.

2. p. 112

3. p.17

4. p. 13

5. Vol. i. no. 1, p. 16.

6. p. 5, n. 5.

7. *Philosophical Works*, ed. J. M. Robertson, 1905, p. 121. Although written English was a language of phonograms in the seventeenth century (N.B. how in his *Essay*, e.g., Locke treats words as sounds but not as marks, III, ix, 2, IV, viii, 13, xi, 7, xxi, 4, etc.) and had not yet become the ideographic language into which the standardisation of its spelling later transformed it, I have sometimes modernised the spelling of my quotations so far as this facilitates understanding them to-day.

8. *Ibid.* p. 522.

9. Letter to Mersenne of 20 November, 1629, Oeuvres, ed. C. Adam and P. Tannery (1897), vol. i, p. 76.

10. *Opuscules et Fragments Inédits de Leibniz*, ed. L. Couturat (1903), p. 27.

11. G. Burnet, *Life of William Bedell* (1688), p. 79.

12. J. Wilkina, *Essay towards a Real Character and a Philosophical Language* (1668), p. 13.

13. 0. Funke, *Zum Weltsprachenproblem in England in 17 Jahrhundert*, 1929, p. 53.

14. *A century of the Names and Scantlings of such inventions as at present I can call to mind to have tried and perfected*, 1663.

15. Adam-Tannery, *op. cit.*, vol. i, p. 572.

16. *Cf.* D. Abercrommbie, "Forgotten Phoneticians", pp. 3 ff., in *Transactions of the Philological Society* (1948). I am indebted to Mr. Abercrombie for drawing my attention to Lodwick. Some further references to seventeenth century English discussions of the project are to be found in an article by R. F. Jones in *Journal of English and Germanic Philology*, xxxi (1932), p. 315 ff.

17. Mentioned in a broadsheet of G. Dalgarno's, in MS. Sloane 4377 in the British Museum. *Cf.* J. A. Comenius, *A Pattern of Universal Knowledge* (1651), tr. J. Collier.

18. In his notebook, MS. Sloane 897.

19. In *Eskybalauron*, 1652. Republished with slight alterations in *Logopandecteision*, 1653.

20. L. Couturat, *La Logique de Leibniz* (1907), p. 51. I have been unable to locate a copy of this scheme.

21. *Vindiciae Academiarum* (1654), p. 21.

22. 0. Funke, *op. cit.*, pp. 37 ff. Cf. Dugald Stewart's remarks quoted in the Maitland Club's edition of Dalgarno's works (1834), pp. v ff.

23. Cf. MS. Sloane 4377.

24. *Cf.* Ms. Sloane 4377.

25. *Mercury, or the Secret, Swift Messenger*, pp. 105 ff.

26. *Leviathan*, pt. iv, ch. 46. *Cf.* Leibniz, *Philosophischen Schriften*, ed. Gerhardt (1865-90), vol. iv, p. 145.

27. Entry in *Journal Book of the Royal Society* under May 14, 1668. There is no trace of any report by this committee in the Society's *Register Book* or in its library catalogue.

28. *Op. cit.*, vol. vii, p. 7.

29. *Op. cit.*, p. 76 ; cf. p. 60.

30. *Ars Magna Sciendi*, Amsterdam (1669). Cf. Hobbes, *De Corpore*, I, i, 2. Bacon had noted in his *Novum Organum*, II, xxvii (*Philosophical Works*, ed. Robertson, p. 335) that "The mathematical postulate that if two things are equal to the same thing they are equal to one another is conformable with the rule of the syllogism in logic which unites propositions agreeing in a middle term." Cf. Locke, *Essay*, IV, xii, 15 for a speculation on the subject.

31. *Eg.* in a letter to Oldenburgh, op. cit., vol. vii, pp. 11 ff.

32. *De Augmentis Scientiarum*, bk. vi, ch. ii (op. cit., p. 533).

33. *The Tongues of Man* (1937), p. 70.

34. *Logical Syntax of Language*, tr. A. Smeaton (1937), p. 7, and *Introduction to Semantics* (1942), p. 1.

35. *The Axiomatic Method in Biology* (1937).

36. E.g. W. Ostwald in *International Language and Science* (1910), p. 55, and J. H. Woodger in *International Encyclopaedia of Unified Science*, vol. ii, no. 5, pp. 1 ff.

37. E.g. B. Russell, *Introduction to Mathematical Philosophy* (1920), p. 183, but cf. "Logical Atomism", in *Contemporary British Philosophy*, First Series (1924), p. 377 ; cf. also J. R. Weinberg, *Logical Positivism* (1936), p. 25.

2

ON A CONCEPT OF DEGREE OF GRAMMATICALNESS

The object of this paper is to clarify a concept of degree of grammaticalness that has appeared occasionally in the recent literature of generative grammar. I aim to make two points. First, the concept of degree of grammaticalness may be distinguished from any semantic concept of intelligibility. Second, it may be explicated in the light of a certain kind of confirmation-theory, the defining conditions for which turn out to explain the non-statistical character of generative grammars.

I

In his seminal work *Syntactic Structures* (p. 35 f.) Chomsky pointed out that a sentence like

John enjoyed and my friend liked the play

was "much less natural", as he called it, than the alternative

John enjoyed the play and my friend liked it.

Sentences of the former kind are generally marked, he claimed, by special phonemic features such as extra long pauses, contrastive stress and intonation, failure to reduce vowels and drop final consonants in rapid speech, etc. These features normally mark the reading of non-grammatical strings. "The most reasonable way to describe this situation would seem to be by a description of the following kind: to form fully grammatical sentences by conjunction, it is necessary to conjoin single constituents; if we conjoin pairs of constituents, and these are major constituents (i.e. "high up" in the [phrase-structure derivation] diagram) the resulting sentences are semi-grammatical; the more completely we violate constituent structure by conjunction, the less grammatical is the resulting

sentence." Similarly, in a later chapter (p.78) Chomsky remarked that any grammar which distinguished abstract from proper nouns would be subtle enough to characterise the difference between two groups of sentences he listed, and he claimed that there was a clear sense in which the sentences of one group were more grammatical than those of the other.

Subsequently many critics of Chomsky, e.g. Jakobson[1], remarked that sentences which he regarded at best as semigrammatical could in fact be given a meaning. The famous string

Colourless green ideas sleep furiously

was even incorporated into a poem. Also Hill's results[2] suggested the unprofitableness of any attempt to construct, or appraise, a scale of grammaticalness on the basis of informants' idiosyncratic acceptances or non-acceptances of individual strings of words. To these critics Chomsky replied[3] by granting them their premisses and denying their conclusion. Certainly, when given a grammatically deviant utterance we attempt to impose an interpretation on it, exploiting whatever features of grammatical structure it preserves and whatever analogies we can construct with perfectly well-formed utterances. But we must distinguish, said Chomsky, between a class of utterances that need no analogic or imposed interpretation and others that can receive an interpretation by virtue of their relations to properly selected members of this class. And Chomsky then proceeded to outline a measure of degrees of grammaticalness in terms of a hierarchy of the categories appearing in the rules of a generative grammar[4]. Katz[5] now claims that this measure is unsatisfactory. He points out that there are some pairs of strings [X, Y] such that Y is much more intelligible than X even though according to Chomsky's example of an n-level hierarchy of categories X would have a higher degree of grammaticalness than Y. E.g. for X we may take

(1) The beef cut sincerity

and for Y

If there is any truth in what he says, it would be to insist foolish.

And Katz argues that indefinitely many counter-examples of this kind could be constructed, since for Chomsky a string fails to receive a representation at level i (and at all higher levels) if it violates even a single restriction holding between a pair of categories at this level. If of two strings, S and S′, S violates just one restriction which prevents it from reaching i-level grammaticalness, while S′ succeeds in reaching i-level grammaticalness but nevertheless has a great deal of structural distortion, then S may well be more intelligible than S′.

But the trouble is that Chomsky applied non-semantic criteria (e.g. phonemic features) in establishing the data about the greater or less grammaticalness that were to be explained by this theory of degrees of grammaticalness. Katz, on the other hand, is constructing his counter-examples with the aid of a semantic criterion – greater or less intelligibility. On the face of it, therefore, Chomsky's problem was not Katz's, and Katz's criticisms of Chomsky appear at first sight to arise from *ignoratio elenchi.* Perhaps it will be objected that Chomsky's problem ought to have been Katz's, and Katz's criticisms are thus justifiable. But I shall now argue that Katz's solution of his own problem shows up the difficulty of identifying degree of grammaticalness with degree of intelligibility.

Katz sets out to explicate the concept of a semi-sentence, where a string is said to be a semi-sentence of the language L if and only if it is not generated by an optimal grammar of L and has sufficient structure to be understood by the speakers of L. He claims that a speaker knows (in the sense in which he knows the rules of the grammar of his language) a system of rules – transfer rules – that enables him to associate a non-null set – comprehension set – of grammatical sentences with each semi-sentence. This association is performed on the basis of the structure that the semi-sentence has; and the speaker's understanding of the semi-sentence is nothing other than his understanding of the sentence in the set with which the semi-sentence is associated. So that a semi-sentence is ambiguous n ways if and only if its comprehension set includes at least one subset of n sentences, none of which is a paraphrase of any other, and includes no subset of more than n such paraphrase-independent sentences. Katz concludes that a theory of semi-sentences cannot be solely a syntactic theory but must contain a semantic component rich enough to provide some means of deciding when two sentences are paraphrases of each other.

But what role does this measure of ambiguity have in Katz's theory of semi-sentences? Katz does not explicitly state that it is to be regarded as a measure for degree of grammaticalness. So let us suppose that it is not to be so regarded. Then Katz's theory suffers from two defects. First, it treats the concept of semi-sentencehood qualitatively, and not comparatively or quantitatively. It has no means of distinguishing between semi-sentences in respect of their degree of failure to be fully grammatical. Secondly, it introduces the concept of a comprehension-set with varying numbers of paraphrase-independent members, and yet assigns no relevant function to this variation other than as a measure of ambiguity.

Let us therefore suppose instead that this measure of a semi-sentence's ambiguity is regarded by Katz as a measure of its intelligibility and thus, for him, of its degree of grammaticalness. But there are several difficulties here too.

First, there seem to be some direct counter-examples, e.g. our intuitions about semi-grammaticalness would normally lead us to suppose that

(2) Three men has left

and

(3) The men has left

have precisely the same degree of grammatical deviance, or conceivably we might suppose that (2) is worse than (3) because the error of putting a singular verb-form with a plural noun-form is made even more glaring by the presence of a plural numeral as well. Yet on the proposed interpretation of Katz's theory (3) is less grammatical than (2) because, while (3) is ambiguous between

The men have left

and

The man has left,

(2) is scarcely ambiguous at all.

Secondly, ambiguity exists even among fully grammatical strings. Indeed some fully grammatical strings are more ambiguous than some semi-sentences. E.g.

(4) He looked over his desk and set out his case

is more ambiguous than (3) and much more so than (2). It seems intuitively objectionable to select as a measure of semi-sentences' degree of grammaticalness a feature that some, but not all, semi-sentences share with some, but not all, full sentences.

Thirdly, even as a measure of intelligibility Katz's theory is not wholly satisfactory. The extent of a semi-sentence's ambiguity is equally affected, on his theory, by any difference of meaning whatever that prevents one sentence in the semi-sentence's comprehension-set from paraphrasing another. But it is easy to find examples that make this egalitarian attitude towards paraphrase-breakdowns seem implausible. E.g. take the semi-sentence

(5) Man bit dog

It has a comprehension-set that contains at least

The man bit the dog

A man bit a dog

The man bit a dog

A man bit the dog

Now, *pace* Katz[6], none of these sentences are exact paraphrases of one another. But the differences of meaning between them are so slight that some speakers of English might be tempted, like Katz, to treat them as paraphrases of one another. Consider, on the other hand, a semi-sentence like

(6) A man a dog bit

It is not unreasonable to assume in this context that the comprehension-set for (6) need contain no other sentences than

A man bit a dog

A dog bit a man.

Yet the difference of meaning between these two sentences seems intuitively greater than that between any of the sentences contained in the comprehension-set for (5). Correspondingly one would suppose (6) to be less intelligible than (5), though Katz's theory implies the opposite. It looks as though any measure of semi-sentences' intelligibility that appealed to a comprehension-set would have to pay regard to the semantic distance, as it were, between the paraphrase-independent members of this set as well as merely to the number of these members.

Perhaps, however, it will be urged that semi-sentences' degree of grammaticalness might still be identifiable with their degree of intelligibility, provided that the latter be measured in some way that avoids the difficulties encountered by Katz's measure. Even Chomsky[7], it may be pointed out, intended his procedure for projecting a hierarchy of categories from a grammar to be construed as part of the explanation of how speakers acquire the ability to comprehend deviant strings. If such a projection-procedure is a component in a language-learning device, then according to Chomsky that device acquires the automatic ability to comprehend deviant strings – deviant in the sense determined by the hierarchy of categories – when it learns the grammar of a language. But Chomsky intended his procedure only as a part of the explanation of how speakers

acquire the ability to comprehend deviant strings, and certainly one can at least conceive that lexical considerations, as well as grammatical ones, may affect degree of intelligibility. Indeed the conventional thing to say about (1) is not that it is ungrammatical but that the normal meanings of the words don't allow the sentence to make sense.

No doubt there is another way of looking at (1). We can suppose our grammar so detailed in the distinctions it makes between different categories of nouns, adjectives, etc. that it is capable of producing rules which show (1) to be grammatically (rather than lexically) deviant. But it has yet to be shown exactly what theoretical or technological purposes are best served by such an extension of grammar far beyond its conventional field of study; the explanation of native language learning by normal children? the education of mentally-retarded children? the teaching of foreign languages? the treatment of aphasia? machine translation? Until we are sure that there is no purpose for which conventional limitations on the scope of grammatical enquiry are appropriate, we should do well not to close the door on the possibility that the relative unintelligibility of certain strings may best be given a lexical rather than a grammatical explanation and these explanations may be of very different kinds.

It seems to me that any proposal to measure degree of grammaticalness by degree of intelligibility does seek to close this door. It assumes relative intelligibility to be explicable solely by reference to grammatical structure[8], and postulates an ordering of the latter that is identifiable with an ordering of the former. But what we need instead are at least two scales, not one. We need one or more non-semantically calibrated scales of grammaticalness, and, independently, one or more scales of intelligibility. We might then be able to assess the extent to which degree of intelligibility is a function of degree of grammaticalness, according to one or more ways of measuring these degrees. And we might thereby be helped to estimate the utility of extending our grammatical analyses right up to the level at which we show (1) to be grammatically deviant. For example, suppose degree of intelligibility were found to correlate with degree of grammaticalness up to a certain level of detail in grammatical analysis, and suppose this correlation began to alter when rules were added to the grammar that imposed yet further restrictions on the generation of grammatical sentences. One might wish to take this as an indication that in some fields (dependent in part on the measures of grammaticalness and intelligibility employed) the likelihood of intelligibility is maximised by transferring the balance of attention at a certain point from grammar to lexis. I would be inclined to conjecture that this may often occur in learning to speak, or translate from, a foreign

language which belongs to the same general culture as one's own. In learning the lexical equivalences *the/le, beef/bœuf, cut/couper, sincerity/sincérité*, a Frenchman has already learnt enough to know that (1) is relatively unintelligible. If there are grammatical rules of English that are responsible for this unintelligibility he has learnt them implicitly through learning the lexical equivalences. But if no special grammatical rules are needed here – whether for the purposes of pedagogy or of translation-programming – then in this context at least the unintelligibility of (1) is not to be traced to grammatical deviance.[9]

II

Perhaps a defender of Katz's approach would rejoin to my criticisms that our intuitions in regard to semi-sentences' degrees of grammaticalness cannot in practice be distinguished over a sufficiently wide range of variation from our intuitions in regard to intelligibility. "The project of constructing a wholly non-semantic measure of grammaticalness", he might argue, "is a vain one because it corresponds to no clearly distinguishable reality in the total linguistic situation. Construct a Chomsky-type hierarchy of categories, if you like. But, in the face of Katz's semantically based counter-examples, what reason is there for taking such a hierarchy to measure anything of importance at all, let alone to measure degree of grammaticalness?"

I want to argue that there is a good reason. The reason is that the Chomsky-type degree of grammaticalness of a string, qua possessing a certain structure, may be identified with the highest degree of confirmation, i.e, evidential support, that is obtainable, on a suitable selection of evidence, for the statement that this string, qua possessing this structure, is fully grammatical. In order to show that such an identification should be acceptable I shall first examine some of the properties of confirmation-functors that can measure degree of evidential support for theses of generative grammar, and then indicate how such functors can come to measure degree of grammaticalness.

We often talk of the extent to which a child has learned certain parts of its native grammar or the extent to which a student has learned certain parts of a foreign grammar. Of course there is no universally accepted system of marking or grading examination scripts in schools and universities. But every school and university has some such system, and certain features are common to most systems. In particular, in testing grammatical knowledge we normally judge an examinee's merit by the variety of non-deviant sentences uttered. The examinee who makes two different kinds of grammatical mistake in a prose composition suggests that he knows less grammar than the examinee who merely makes two

mistakes of the same kind, and the former would normally be marked or regarded lower than the latter if the three kinds of mistake are all equally serious. Similarly the recent literature of linguistics has contained many comparative judgements about the merits of different generative grammars produced by investigators of various languages, and these merits are alleged to be shown at least in part (though other factors, such as relative simplicity, or suitability as a foundation for semantics, may also be taken into account) by the variety of hazards that the proposed rules surmount in avoiding the generation of grammatically dubious sentences. What seems to be implicitly employed in these judgements is a two-place functor c[H, E], where H is a hypothesis asserting that any string generated by such-and-such a set of rules (the rules implicitly employed by a student or explicitly stated by a linguistic investigator) is a grammatical sentence, and E reports one or more intuitions about the grammaticalness of individual strings thus generated. We may assume these functors to denote real numbers ≥ 0, so that we can either assign numerical values, in the form c[H, E] $= n$, or at least draw comparisons of the form c[H, E] $>$ c[H′, E], which assert that H is better supported by E than H′ is.

What other assumptions are plausible? Presumably (call it the *Equivalence Condition*) if H is logically equivalent to H′ and E to E′ then c[H, E] = c[H′, E′]. Presumably also (call it the *Instantial Comparability Condition*) if S is a substitution-instance of the hypothesis H (asserting the grammaticalness of a particular string if generated by the rules described in H), and if S does not mention any element of H's domain – i.e. any string – that is mentioned in E, and if S′ is analogously related to H′ and E′, then c[H, E] $>$ c[H′, E′] if and only if c[S, E] $>$ c[S′, E′]. I.e. the relative weight of evidential support for an individual sentence's grammaticalness, in virtue of its having such-or-such a structure, varies directly with that for the relevant rule or rules and vice versa. For example, Chomsky[10] in effect agrees with Hill[11] that in generative grammar the grammaticalness of a string is to be judged only in the light of its supposed method of generation. A string is otherwise insufficiently identified for its grammaticalness to be judged. Conversely support for a string so identified is also support for the generative rules that help to identify it.

Now it can be demonstrated[12] from these two conditions that (*Instantial Conjunction Condition*), where S^1, S^2, $\ldots S^n$ are all substitution-instances of H mentioning no element of H′s domain – i.e, no string – that is mentioned in E, then, so long as we may assume that at least n applications of the rules described in H have in fact been made, c[S^1, E] = c[S^1 & S^2 $\ldots$& S^n, E]. I.e. two or more applications of a set of

generative rules are normally just as well supported by relevant evidence as a single application. The importance of this Instantial Conjunction Condition lies in the fact that, where we can have c[X & Y, E] = c[X, E], there c[X, E] cannot, except in certain limiting cases, be either a probability or a function of probabilities. This is obvious enough in the case of simple probabilities, but it can also be demonstrated[13] for any function of probabilities. The resistance to statistical support that Chomsky has observed[14] in regard to assessments of grammaticalness is thus a necessary feature of any confirmation-functor that satisfies two formal or quasi-formal conditions, viz. the Equivalence and Instantial Comparability Conditions. For these conditions determine certain properties of c[S, E] irrespective of the subject-matter with which S and E are concerned. Of course, the Instantial Comparability Condition is only of permanent interest in relation to universal hypotheses with an infinite domain. For, if the domain of H were finite, there might come a time when we had so much evidence available about the elements of that domain that, if we were to cite all available evidence in E, (in accordance with what Carnap[15] called the Requirement of Total Evidence), we could no longer find a substitution-instance, S, of H satisfying the requirement that S does not mention any element of H's domain which is mentioned in E. But this does not constitute a reason for not applying the Instantial Comparability Condition to the special case of generative grammars, since the rules of a generative grammar are supposed to be capable of generating an infinite number of sentences.

A more general result is also available, if three other conditions may be assumed.

First, if a syntactical theory generates indisputably ungrammatical sentences it needs to be rejected and replaced by a theory that differs from it in some respects. Hence if the evidence for a hypothesis of the kind we have been considering includes mention of a clear counter-instance the evidence, gives no support to the hypothesis as it stands. Contrapositively, if the hypothesis is supported, its negation is not. We therefore assume (Negation Condition) that if c[H, E]>0, then c[~H, E]=0. Our confirmation-functors thus measure degree of support only, not degree of undermining also: we do not have, e.g., c[H, E]=1–c[~H, E]. We are out to weigh just the typical build-up by eliminative induction – the accumulation of varied unsuccessful attempts to falsify a universal hypothesis.

Secondly, we must assume that any evidential support a universal theory has is passed on to its logical consequences. More specifically (*Consequence Condition*), for all E, H and H′, if H is logically-deducible from H′, then c[H, E] $\geq$ c[H′, E]. There would be little point in scien-

tific generalisations if the applications of a universal hypothesis H to particular cases were not at least as well supported as H itself by the evidence for H: our general knowledge would be useless. And scientific system-building – the axiomatisation of a body of universal hypotheses – would be equally useless if evidential support were not automatically passed on by a conjunction of universal statements to any other universal statement logically deducible from it.

Thirdly, any hierarchy of grammatical categories such as is described by Chomsky may be regarded as presenting a series of tests that are applicable to hypotheses about rules of generative grammar. At each level of the hierarchy we are presented, in ascending order of subtlety, with distinctions between two or more categories of expression (e.g. verb/noun, transitive verb/intransitive verb, animate noun/inanimate noun) by reference to which we may vary the circumstances of the test we apply. (E.g. where a rule is formulated in terms of any verb and any noun one test will consist in examining it in relation to one sentence that contains a transitive verb and an animate noun at the relevant points, another that contains an intransitive verb and an inanimate noun, and so on.) Thus there may be assumed to be (*Evidential Ordering Condition*) a fixed serial order for the application of tests to hypotheses, so that if C is the class of tests to which one hypothesis has been subjected and C′ is the class to which another has been subjected, either C is included in C′ or C′ in C.

Now from the Negation, Consequence and Evidential Ordering Conditions it is easily demonstrated that if $c[H', E] \geq c[H, E]$ then $c[H \,\&\, H', E] = c[H, E]$, where H and H′ are universal hypotheses. We thus have the anti-probabilistic situation, $c[X \,\&\, Y, E] = c[X, E]$, again. Confirmation-functors for generative grammars normally resist interpretation into the symbolism of the probability-calculus, whether they be concerned with the application (substitution-instances) of theories or with the theories themselves.[16]

I wish now to argue that to the extent we can understand such a concept of degree of evidential support for generative grammar we must also be capable of understanding a wholly non-semantical concept of degree of grammaticalness. For to discover the degree of grammaticalness of a string, qua possessing a certain structure, we simply need to determine how well supported, on a legitimate selection of the available evidence, is the thesis that any string possessing this structure is grammatical. Of course at any level a test may produce a counter-instance which belongs essentially not to that level but to some higher one. Our test may appear to fail at the verb/noun level, e.g., not because of the verb formulation as such, but because we happen to have chosen an

intransitive verb. If a formulation with some transitive verb is fully grammatical then the failure belongs not to the verb/noun level, where we can afford to neglect the counter-instance, but either to the transitive/intransitive level or to some yet higher one. In short, the higher up our categorial hierarchy we can go, i.e. the more refined the distinctions we can draw while still obtaining at least same indisputably grammatical sentences from our generative hypotheses, the better established is our thesis on the evidence thus far available, and, correspondingly, the more grammatical is the given string qua possessing the structure in question. Thus, since no-one can investigate generative syntax rationally without making some comparative (and even quantitative) estimates of such degrees of confirmation on the basis of purely qualitative intuitions about grammaticalness or non-grammaticalness in particular cases, we have a built-in guarantee that so far as syntactical studies are distinguishable from semantic ones we have intuitions that go to establish degrees of grammaticalness which are distinguishable from our intuitions in regard to greater or less intelligibility.

Notes

1. R. Jakobson, "Boas's View of Grammatical Meaning", *American Anthropologist* lxi (1959) memoir no. 89 of Am. Anth. Assoc.

2. A. A. Hill, "Grammaticality", *Word* xvii (1961) p. 1ff.

3. N. Chomsky, "Some Methodological Remarks on Generative Grammar", *Word* xvii (1961) p. 219 ff.

4. *Ibid.*, p.236ff.

5. J.J. Katz, "Semi-sentences" in *The Structure of Language*, ed. J. A. Fodor & J. J. Katz (1964) p. 400 ff.

6. *Ibid.*, p.411.

7. *Ibid.*, p.404.

8. In his concluding remarks (*ibid.*, p. 416) Katz seems to recognize some of the risks inherent in this assumption.

9. So far as this is so there is something to be said for the familiar philosophical thesis that conceptual principles, or rules of logical grammar, may be distinguished from rules of linguistic syntax by their greater invariance under translation. The extent to which grammatical rules are actually needed (as distinct from being theoretically possible) in the analysis of a language also tends to be reflected in the extent to which grammatical category-names are readily freed from semantic implications: contrast the terms *transitive/intransitive* with *animate/inanimate*. Note also that even in formal-logical theory "x is both square and not-square" is normally not regarded as ill-formed, though it is L-false. Presumably therefore "x is both square and round" and "x is both square and shapeless" are equally to be regarded as well-formed.

10. *Loc. cit.*, p. 228f.

11. *Loc.cit.*, p.9.

12. The demonstration is given in L. Jonathan Cohen, "A Logic for Evidential Support" forthcoming in *British Journal for the Philosophy of Science.* An earlier version of this article was delivered in the form of two lectures to the University of Liege and the Centre National Belge de Recherches de Logique on March 26 and 27, respectively. Cf. also L. Jonathan Cohen, "What has Confirmation to do with Probabilities?" forthcoming in *Mind.*

13. "What has Confirmation to do with Probabilities?", forthcoming in *Mind.*

14. *Syntactic Structures*, p.16f.

15. R. Carnap, *Logical Foundations of Probability* (1950) p. 211.

16. Instead, as it turns out, confirmation-functors of this kind are definable, and their logical properties formally provable, within a generalised modal logic, as is shown in "A Logic for Evidential Support" (forthcoming).

3

THE SEMANTICS OF METAPHOR

Metaphor as a feature of sentence readings

Professor Sadock's position is that metaphor is not a linguistic problem at all, because the mechanisms underlying metaphor exist independently of language. The use of "a lion on a warrior's shield," he says, manifests "the same analogical urge that functions in the issuance and apprehension of verbal metaphors." But it is vital here to distinguish metaphor, on the one hand, from similes and other nonmetaphorical analogies, on the other. There is certainly no special linguistic problem about such explicitly figurative sentences as (1) or (2):

(1) He is as brave as a lion.

(2) He is like a father to her.

The linguistic problem arises instead because of implicit figuration, as in (3) or (4):

(3) He is a lion in battle.

(4) The child is father to the man.

The problem is to explain how such sentences can have the meaning they do, when they are false or nonsensical if taken literally. In other words, the analogical urge operates sometimes in literal speech, sometimes in metaphorical It can produce both and so cannot explain the difference between the two.

My own position, as against Sadock's, is that the fundamental problem about metaphor is a problem for our theory of *langue*, not for our

theory of *parole*. But I speak here only about metaphor, and not about other figures such as irony, litotes, allegory, or simile.

Various considerations operate in favour of this position, although they neither severally nor jointly entail it.

First, if the synchronic description of a language takes no account of metaphor, it provides an inadequate basis for diachronic explanations of semantic innovation. Language is full of dead metaphors like "inflamed passion," "feeble argument", or "rain of blows": How have they got there?

Again, it is clearly characteristic of certain categories of artificial languages that they must lack any possibility of metaphor. Programming languages for computers, like Fortran, or interpreted formal systems, like Carnap's, would be very seriously flawed in the performance of the tasks for which they are severally designed if they allowed their component words or symbols to be attributed new and unstipulated meanings in certain contexts. It is arguable, therefore, that we radically blur the difference between these kinds of artificial languages on the one side, and ordinary natural languages, like English (or artificial languages for everyday use, like Esperanto), on the other, if we do not allow essentially for the possibility of metaphor in our analysis of the latter. For example, it is characteristic of a natural language sentence like (5)

(5) He is a lion

that it is indefinitely rich in possible meaning. Alongside one or more literal readings it admits also an indefinite variety of metaphorical ones. A theory that does not allow this will never satisfy the intuitions of people who are expert in the use of language, like the best creative writers. (The French classical tradition was an exception, but a self-conscious and theoretically motivated one.)

Thirdly, metaphor cannot be explained within a theory of speech acts, because a supposed speech act of metaphorizing would differ from standard types of speech act in an important respect. If Tom utters the sentence

(6) I am sorry,

he may well be apologizing. But, if I utter the sentence

(7) Tom said that he was sorry,

I am not apologizing myself; I am just reporting Tom's apology. The original speech act is overridden by the passage from oratio recta to oratio obliqua. Now metaphor behaves quite differently. When Tom describes his friend by saying

(8) The boy next door is a ball of fire,

Tom's description can be fully understood only by someone who understands the metaphor. But equally, if I myself report later

(9) Tom said that the boy next door is a ball of fire,

my report of Tom's utterance can also be fully understood only by someone who understands the metaphor. The metaphor is not overridden by the passage from oratio recta to oratio obliqua: the oratio obliqua sentence contains the same element of metaphorical meaning that the oratio recta contained. Arguably, therefore, metaphorical meaning inheres in sentences, not just in speech acts. We should be perfectly happy to say here, "What Tom said is true," not just, "What Tom meant is true."

This point is a very serious difficulty for anyone, like Searle (cf. his paper in the present volume) who wants to construe metaphor solely in terms of speaker's meaning – the meaning of the utterance rather than of the sentence uttered. The metaphoricalness, or special character, of (8) is preserved under transformation into indirect discourse: the apology-making character of (6) is not. Why is this? The simplest explanation seems to be that metaphoricalness is a property of sentences. Those, like Searle, who wish to reject such an explanation bear the onus of presenting a convincing alternative one. Otherwise their theory limps.

No doubt the linguistic context of utterance often fails to provide sufficient cues for us to disambiguate a metaphorical sentence. We must turn then to whatever features of the nonlinguistic setting are relevant. But just the same is true of nonmetaphorical sentences, as with the meaning of "unsatisfactory," say, in

(10) He's unsatisfactory.

It would be a mistake to suppose that this familiar component of communicational situations provides any special support for a speech-act theory of metaphor.

Fourthly, if our linguistic theory gives no account of metaphor, it may well make so much the poorer a contribution to the framework within which psycholinguistic investigations are carried on. Language-learning infants often speak in metaphors without knowing the literal meanings of the sentences they utter (Cohen and Margalit, 1972, p. 470).

But there is another reason why the stock attitude of linguistic theorists to metaphor is incorrect; and this reason is an intrinsically stronger one, because it looks inwardly to the structure of synchronic semantics, rather than outwardly to the use of semantic analyses in etymological explanations, in comparisons between natural and artifical languages, in

speech-act theory, or in psychological inquiry. I have in mind a certain choice that exists in the methodology of natural language semantics, which I shall refer to as the choice between the method of cancellation and the method of multiplication. This choice arises in relation to quite a range of other problems besides that of metaphor; and my argument will be that whatever choice of method is made to deal with metaphor both reinforces, and is reinforced by, the choice of the same method in the other cases. In particular, if we need to introduce the method of cancellation into our semantics in order to deal with certain kinds of nonmetaphorical sentences, we might just as well exploit its potential for the analysis of metaphor.

Let us turn briefly from metaphor to examine these other, non-metaphorical cases. Consider the differences between

(11) It wasn't an insult because it was not intended as such

and

(12) It was an unintentional insult.

Clearly the word "insult" might be described by a compositional semantics as occurring in a sense that is +INTENTIONAL in (11), but not in (12). But such a description chooses what I have called the method of cancellation. It supposes that the occurrence of the adjective "unintentional" in (12) has cancelled the +INTENTIONAL feature in this occurrence of the word "insult," (because the retention of that feature would render the sentence self-contradictory). We need then, for the description of (12) and (12), just one lexical entry for "insult," alongside some general set of rules for semantic feature-cancellation in the process of composing sentential meaning. Correspondingly both (11) and (12) are unambiguous.

Alternatively, however, we could employ the method of multiplication. We could have two lexical entries: "insult$_1$" which is +INTENTIONAL, and "insult$_2$" which is −INTENTIONAL. We would then dispense here with the use of any rules for semantic cancellation, and both (11) and (12) would now be ambiguous. One of the two underlying sentences for (11) would have the same meaning as on our previous interpretation of (11), the other would be rather a non sequitur. Similarly, one of the two underlying sentences for (12) would be self-contradictory, the other would have the same meaning as our previous interpretation. But what is crucial here is that the alternative to the method of cancellation is a method that doubles the number of relevant lexical entries.

Perhaps I should add here that there is not a third – quasi-Gricean – possibility. It will not do to say that "insult" never has the feature

+INTENTIONAL, but carries with it, in normal contexts, a conversational implicature of intentionality, which happens to be cancelled in (12). The reason why this possibility is not available is because it cannot explain the force of the "because ..." clause in (11). (In regard to the logical particles of natural language, Grice's theory of implicatures does afford an alternative approach to the method of cancellation, but arguably an inferior one. See Cohen, 1971; Cohen, 1977a[1].)

It would be easy to give many more examples of the issue about +INTENTIONAL, which is a particularly familiar hazard in the register of jurisprudence. But other kinds of example are also frequent. Consider

(13) I tried to warn him but he didn't hear me

and

(14) I warned him, but he didn't hear me.

The method of cancellation would describe the word "warn" as occurring in a sense that is +UPTAKE in (13) but not in (14). But the method of multiplication would generate two correspondingly different lexical entries for the word "warn."

Another familiar feature in relation to which we are faced with a choice between the method of cancellation and the method of multiplication is the feature +ANIMATE. Consider the difference between

(15) A lion eats ten pounds of meat a day

and

(16) A stone lion needs no feeding.

According to the method of cancellation, the feature +ANIMATE is present in the occurrence of "lion" in (15), but is cancelled in (16): according to the method of multiplication the lexeme "lion_1' occurs in (15) and a different lexeme, "lion_2' occurs in (16). A vast number of analogous examples can be given, like "china dog," "plastic flower," and so on. The method of cancellation avoids the need to multiply lexical entries in order to deal with such phrases. But the method of multiplication promotes a second lexical entry, with the sense "replica of a lion," "replica of a dog," "replica of a flower." Of course, one could hope to avoid cancelling or multiplying anything in relation to "lion" if one supposed that "stone" in (16) meant "stone replica of a" But the focus of methodological choice – choice between the method of cancellation and the method of multiplication – has now just shifted to "stone." The contrast is now between the occurrence of "stone" in (16) and its occurrence in

(17) A stone wall surrounds the garden.

Again the feature for mode of functioning, or mode of operation, would be at issue for a word like "car" in such sentence pairs as

(18) We will pack the luggage into the car

and

(19) We will pack the car into the luggage.

The method of cancellation allows us to suppose the same lexeme occurs in the composition of both (18) and (19). The method of multiplication requires us to suppose two lexical entries in relation to every inanimate object as well as in relation to every plant or animal: one for the name of the original, and a second for the name of its replica.

If we are always to prefer the method of multiplication wherever possible, our linguistic theory can hardly have anything of interest to say about metaphor. Consider, for example

(20) Their legislative program is a rocket to the moon.

The method of multiplication can either treat (20) as involving some neologism that could, if it becomes established in the language, figure as "rocket$_2$' in a listing of English lexemes. Metaphor is then reduced to polysemy. Or alternatively the method of multiplication could reduce metaphor to simile by adding to the ordinary copula a second lexeme "is$_2$," in the sense of "is like," and making similar additions for other predicative morphemes. In neither case will metaphor appear as anything other than a rather peripheral or superficial item in the immense motley of lexical variety. Its existence will have no particularly close dependence on certain special modes of sentential composition, and those linguists who have postponed its treatment until they have an adequate semantics for nonmetaphorical sentences will be immune to any charge of methodological hysteron proteron.

But is it right to adopt the method of multiplication in cases like (12), (14), (16), and (19)? Scientific inquiry tends to make more progress if it avoids unnecessary multiplications of entities, and it is this Ockhamian policy that the method of cancellation pursues. If we could formulate principles controlling feature-cancellation in the composition of sentence meanings we should gain a powerful, general insight into the semantics of natural language, which will be forever denied us if we are content to follow the method of multiplication and merely list or catalogue its proliferation of lexical entries. No doubt there are some points at which the method of cancellation ceases to pay dividends. Where the meanings

of a polyseme have grown apart and taken on different positive features, like "funny" meaning "comic" and "funny" meaning "strange," the best procedure may well be to follow the method of multiplication and treat the word semantically just as if it were a homonym. Where a particular combination of words, like "kick the bucket," has lost all touch with the component words' meanings in other combinations, the phrase may need to be listed as an idiom alongside single words and have its own lexical entry. But the fact that in such circumstances we are forced into falling back on the method of multiplication constitutes no argument against pursuing, wherever possible, the economies and insights that are afforded by the method of cancellation. The method of multiplication merely records what meanings exist, without explaining how they are generated. It lists elements but does not explore relations.

1. The problem of feature cancellation

What then should be said about metaphor in a componential semantics if our policy in analysing the composition of sentence meanings is to be the Ockhamian one of preferring, wherever possible, the method of cancellation to the method of multiplication? Clearly we have to regard metaphorical meaning as being put together by a different kind of cancellation from that which appears in nonmetaphorical cases like (12), (14), (16), and (19). But what is the nature of the difference?

Let us distinguish broadly between those semantical features that represent attributes which are empirical, immediately evident, or relatively obvious, and those which are inferential, intellectually appreciated, or relatively latent. It will be convenient to refer to these groups of features as "empirical" and "inferential," respectively, though determinate classification will often be difficult and there are bound to be many borderline issues. Then a tempting hypothesis is that in the normal, literal cases it tends to be inferential features (though not necessarily all of them) that are cancelled, whereas in metaphorical sentences it is empirical features (though again not necessarily all of them) that tend to be cancelled. Thus in the examples considered earlier the cancelled features +INTENTIONAL, +ANIMATE and +UPTAKE are all inferential ones. But in the metaphorical sentence (20), the legislative program is presumably said to be a rocket only in a sense that has shed such empirical features as +MATERIAL, +AIR-CLEAVING, +CYLINDRICAL, and so on, which are incompatible with the features of "legislative program," while retaining such features as +FAST-MOVING, +FAR-AIMING.

Such a hypothesis seems to jibe well with the common use of metaphor to enliven discussion of an abstract subject with concrete imagery. And

it would be no objection to this hypothesis that even the uncancelled features in a word's metaphorical occurrence may also sometimes be empirical ones, as in

(21) Jane has the face of a wild rose

or

(22) The clouds are made of pure gold.

For, though a woman's face may share the colour of a wild rose blossom, or even have a similar shape, it certainly lacks other empirical properties associated with wild roses, such as their spiky stems; and though sunset clouds share the colour of pure gold, they certainly lack its tangibility.

Nevertheless, there are at least two reasons why this hypothesis is unsatisfactory.

First, even nonmetaphorical cancellation sometimes removes empirical features, as in

(23) They have produced a new breed of animal – a hornless cow.

The feature +HORNED is certainly an empirical one, but it is cancelled from the sense of "cow" in (23) without any intuitable semblance of metaphor. We need a hypothesis that will explain why an empirical description in which one or more of a lexeme's empirical features have been cancelled is metaphorical in (21) and (22) but not in (23).

Secondly, in some cases of metaphorical cancellation there seem to be no empirical features available for removal. Consider, for example,

(24) The troop movements ordered by Wellington were logarithmic in their strategy: by adding complementary units to one another in a particular location, he obtained the same effect as if he had multiplied the number of mutually ill-assorted units in that location.

Here we have an image taken from the highly abstract subject matter of mathematics and applied to the somewhat less abstract topic of military strategy.

It would appear that a more satisfactory hypothesis might be one which exploited the concept of a semantic category. On this view, cancellation would be metaphorical, if and only if, the features cancelled were intrinsic to the lexeme's superordinate semantic category. Thus in (12), (14), (16), and (19), it might be argued, the category remains essentially unaltered. We are still dealing with a name for a type of speech act in (12) as in (11), and in (14) as in (12): we are still dealing with a name for a type of physical object in (16) as in (15), and in (19) as in (18). On the other hand in (20) we are dealing with a name for

something – a legislative program – that does not belong in the same semantic category as "rocket," and in (24) we have a shift from the name of a mathematical operation to the description of a military one. So far so good.

But the trouble here is that what we should need to count as a semantic category for dealing with some sentences cuts across what we should need to count as a semantic category for dealing with others. Thus to maintain the nonmetaphoricalness of (16) we need to think of both lion and stone lion as belonging in a single category, that of PHYSICAL OBJECT. Yet "face" and "rose" also seem to belong in this category, though we should need to think of them as belonging in different categories in order to explain the metaphoricalness of (21) by the proposed hypothesis. If criteria for sameness and difference of semantic category in a natural language have to be systematically adapted to native-speakers' intuitions of metaphoricalness, we can no longer appeal to categorical structure for an explanation of the differences between metaphorical and nonmetaphorical cancellation. The description of categorical structure would be just a restatement of this difference.

I was once inclined therefore (Cohen and Margalit, 1972) to favour a third hypothesis, which relies neither on the "empirical-inferential" distinction nor on the concept of a semantic category. The assumption that lies behind this third hypothesis is that, for theoretical purposes, semantic features should not be classified dichotomously, whether as empirical or inferential, or as categorical or noncategorical, or in any other way, but should rather be arranged in an order of decreasing semantic importance. Then a cancellation would be metaphorical if and only if it removes the most important feature or features. For example, *gold* would be said to have a metaphorical occurrence in (22) because its most important feature +METALLIC would be cancelled.

Now there are several good reasons, quite apart from this issue of metaphorical and nonmetaphorical cancellation, why a semantic theory for natural language should assume the semantic features appropriate in this or that area of a language's vocabulary to be at least partially ordered in importance, rather than merely coordinate with one another. One reason is that we need such an assumption in diachronic semantics. It explains why in borderline cases, where there is a choice between either modifying one word to fit a novel or unusual situation or modifying another, we may be inclined to choose the modification that makes less of an inroad into the core of a word's meaning. For example, the verb naming the action of a hang-glider is "to glide," not "to parachute." A second reason is that to suppose an ordering of this kind provides a certain room for maneuver in the philosophical analysis of a concept.

The claim to be concerned only with the nuclear sense of a concept may provide a valid excuse for not pursuing the analysis of the concept into those thickets of refinement where the argument gets bogged down in dealing with ingenious but farfetched counterexamples. Thirdly, it can be shown (Cohen, 1970, p. 35ff.; Cohen, 1977b, p. 129ff.) that any inductive processing mechanism, like a vocabulary-learning infant, which seeks to learn from the variety of its experiences rather than from the mere multiplicity of experiences of one particular kind, must impose some importance-ordering on the relevant patterns of variation.

However, I do not want to digress here into issues that belong more properly to other inquiries than to linguistic theory. Suffice it to say that the most serious mistake in the Wittgensteinian family-resemblance theory about the meanings of certain descriptive terms is its incompatibility with the assumption of some kind of importance-ordering for semantic features. And any theory of metaphor must operate within the framework of this assumption. Nevertheless, it will not do at all to say, as I once said (Cohen & Margalit, 1972, p. 484), that metaphorical cancellation is distinguished by its always being the cancellation of semantically important features. After all +ANIMATE hardly seems much less important a feature than +METALLIC. Yet its cancellation in (16) does not render the sentence's meaning metaphorical. Moreover, importance is a matter of degree, but though metaphors may be better or worse, a phrase's or sentence's meaning seems to be *either* literal *or* metaphorical.

How in any case is semantical importance to be graded? We must certainly distinguish carefully here between semantical and practical importance–between what is relevant for the applicability of a generic name, and what is relevant for attitudes towards things of the kind named. It may be highly relevant to your mobility whether you have a drivable car or only a toy one, and highly relevant to your safety whether a cow has horns or not. But what counts for semantical importance is the capacity to falsify propositions about the applicability of terms, not the capacity to falsify propositions about what practical course of action is possible or desirable . Hence the features that are relatively more important from a semantical point of view are those that are relatively more distinctive and specific. Because they are less probable (that is, have greater prior improbability in the universe of feature-instantiations), they carry more information. And if that is how semantic importance is to be graded, I suspect that most metaphorical cancellation, like all nonmetaphorical cancellation, begins at the lower end of the scale. Rather general features like +ANIMATE or +METALLIC are highly eligible for cancellation because they are semantically unimpor-

tant; whereas features representing specific peculiarities of appearance or behaviour are considerably less eligible for cancellation because they are semantically much more important.

2. The relevance of the topic-comment distinction

So, though it is quite dear that some metaphorical cancellation occurs at the upper end of the scale, as in, for example,

(25) The poor used to be the blacks of Europe,

the difference between metaphorical and nonmetaphorical cancellation cannot depend on either the nature or the semantic importance of the feature cancelled. But if this difference does not depend on which kinds of elements are connected by the cancellation-relation, and if feature-cancellation itself is a uniform operation, then perhaps we shall get a clue to what the difference does depend on from the direction of the cancellation-relation, as determined by the grammatical functions of the expressions related. If we can distinguish within any phrase or sentence in which cancellation occurs, between the topic-expression – the expression that would normally be isolated by the particle "wa" in Japanese – and the comment-expression, or between the head and the modifier, then metaphorical cancellation seems normally (exceptions are considered below) to be imposed by the topic-expression on the comment-expression, or part of it, or by the head on the modifier, whereas nonmetaphorical cancellation is imposed by the comment-expression on the topic expression, or part of it, or by the modifier on the head. Thus cancellation runs from "legislative program" to "rocket" in (20), from "clouds" to "gold" in (22), from the second sentence in (24) to "logarithmic," and from "Europe" to "blacks" in (25). But it runs from "unintentional" to "insult" in (12), from "but he didn't hear me" to "warned" in (14), from "stone" to "lion" in (16), from "into the luggage" to "car" in (19), and from "hornless" to "cow' in (23).

We have still to explain, however, just why it is that the direction of feature-cancellation normally varies thus from sentence to sentence or sentence-sequence to sentence-sequence. One might perhaps have expected, prior to examining instances, that an operation of this kind would always run in the same direction. Why does that not happen? I suggest that the explanation lies in the overriding linguistic need to make the topic of a sentence or sentence-sequence easily identifiable. Ease of topic-identification must be conserved under variations in the amount of feature-cancellation. If the extent of cancellation (considered as a function of the number and importance of features cancelled) is

relatively small, as in (12), (14), (16), (19) or (23), cancellation runs naturally from comment-expression to topic-expression parallel to the normal direction of modification. But when cancellation is relatively extensive, as in (20), (21), (24), (26), and (27), easy identification of the topic is preserved by the cancellation's running from the topic-expression to the comment-expression; and it is thus, in the grammatically well-formed sentences of a language, that what we commonly call metaphors arise. Of course, a good or apt metaphor will not involve too much cancellation, because then it will have too few points of similarity with that for which it is a metaphor. But if there is too little cancellation in a word or phrase it will not be a metaphor at all. An expression has a metaphorical occurrence in a sentence-sequence if and only if the extent of cancellation (from the expression's semantic features) is so great as to exclude it from identifying a topic independently of some special circumstance (such as an ad hoc convention, a relevant previous sentence in the sequence, a familiar linguistic practice, and so on).

If we consider apparent exceptions to this principle, they will all be found, I suspect, to rely on certain special circumstances for their plausibility. For example,

(26) The brightest star in my sky has just run off with the milkman

is obviously a sentence in which the topic is identified by a metaphorical expression. But this is possible just because the metaphor is a trite one.[2] An original or unfamiliar metaphor, unaided by context, would not identify a topic with any clarity. For example, the sentence "The sparkle on summer dew has just run off with the milkman" is a kind of nonsense if considered as an isolated unit. But its topic would become quite apparent if it occurred in some suitable sequence of sentences that contained

- (27) My wife is the sparkle on summer dew

as an earlier element. Somehow or other, whether by being more readily intelligible, or by being present at least in the first occurrence of metaphor, the comment-making use of a metaphor is primary and any topic-identifying use is secondary and derivative.

A number of amplificatory points need to be made in conclusion.

First, I take it that lexical entries for a natural language can draw no clear distinction between features that are supposed to be "purely linguistic" and features that are supposed to represent common knowledge or commonly accepted beliefs. The *Concise Oxford Dictionary* (6th ed., 1976) tells us quite appropriately that lions are found in Africa and South Asia and that privet is much used for hedging. There is no confusion here between the function of a dictionary and the function of an

encyclopedia. We need encyclopedias to tell us things that are not common knowledge, not things that are. Of course, all a word's semantic features have to be ordered for importance, and the type of feature that we are inclined to call "common knowledge" is normally a relatively unimportant one. Because it is less important it is less insisted upon, and sentences articulating it are more easily treated as synthetic propositions. However, such features frequently provide a vital part of the residual analogy that a metaphorical sentence seeks to convey.

Secondly, I am not claiming that the same order of importance for semantic features extends throughout a language's vocabulary. For example, while the features determining visible appearance are highly relevant for animal names, they are clearly not so relevant for the lexemes that name social roles, such as "policeman," "president," or "banker." The main differences between such lexemes are revealed by inferential rather than empirical features. A plain-clothes policeman is still a policeman, and a water-skiing banker is still a banker.

Thirdly, each of the sentences that I have discussed so far is either obviously literal or obviously metaphorical. Many others are ambiguous between these two possibilities. For example, the sentence

(28) It will never get off its launching pad

has not only a (familiar) literal meaning but also a metaphorical one - in which it might be uttered as a reply to the utterance of (20). The sentences that are normally called metaphorical ones, like (20), (21), (22), (24), (25), (26), or (27), are sentences that appear absurd or fragrantly false if we try to give them a literal reading. But to know how we should understand an utterance of (20) we need to know its topic. Or we could say, in general, that though there is normally a presumption in favour of a literal interpretation, this presumption can be rebutted by suitable contextual cues. Correspondingly in *langue*, if we always consider not just isolated sentences but sufficiently long sentence-sequences, we can take any such sequence to be a literal one throughout, unless this involves an absurdity that would disappear on a metaphorical interpretation.

Fourthly, I have assumed throughout that we cannot dispense with some kind of compositional approach to the semantics of natural language if we are to achieve any determinate progress–anything more than pious generalities–in this area of inquiry. An analytic resolutive methodology, as so many other sciences can testify, leads us naturally into a deeper understanding of richly structured wholes, and the meaning of a natural language sentence, whether literal or metaphorical, is certainly such a whole. But once a compositional approach is accepted we can scarcely avoid being led, for the reasons already canvassed, to prefer

what I have called the method of cancellation to the method of multiplication, to recognize the significance of the direction in which feature-cancellation operates, and to explain the choice of this direction by reference to the extent of the cancellation and the need to conserve ease of topic-identification. I do not underestimate the immense difficulty and complexity of the task that anyone would undertake who sought to construct a rigorously adequate theory of cancellation. But in the present paper I leave been concerned rather to argue the more general issue: How intrinsic to linguistic theory is the problem of metaphor and what, in general, is the best solution for it? I have also not been concerned at all with the corresponding psychological problem: How are metaphors invented and comprehended? But it is reasonable to suppose that until the linguistic problem has a definitive solution, the psychological one cannot be adequately articulated.[3]

Notes

1. The discussion of proposition-compounding particles, like "or" and "if" in these papers could be extended to quantifiers like "some." For example, either "some" has a cancelable semantic feature that implies "not all" or its use normally carries a conversational implicature "not all." A Gricean treatment would begin to run into difficulties with sentences like

   If some of the money has already been found, the rest will be found quite soon.

Note that it would be unreasonable to expect that semantic cancelability will be preserved when such a word is disambiguated by compound phrases like "a part but only a part" and "a part or even the whole,"' respectively.

2. It is tempting to think that some proverbs claim a certain kind of metaphorical truth alongside their literal truth, as in

   A stitch in time saves nine.

If this is the correct way to regard such proverbs, the metaphor will be as trite as the proverb is familiar.

3. I am grateful to Dr. Gillian Cohen for some very helpful comments on an earlier draft of the present paper.

References

Cohen, L. (1971). Some remarks on grice's views about logical particles of natural language. In Bar-Hillel, Y., editor, *Pragmatics of natural language.* Humanities Press, New York.

Cohen, L. (1977a). Can the conversationalist hypothesis be defended? *Philosophical Studies*, 31:pp. 81–90.

Cohen, L. and Margalit, A. (1972). The role of inductive reasoning in the interpretation of metaphor. In Davidson, D. and Harman, G., editors, *Semantics of Natural Language.* Reidel, Dordrecht, The Netherlands.

Cohen, L. J. (1970). *The Implications of Induction.* Methuen.

Cohen, L. J. (1977b). *The Probable and the Provable.* Clarendon Press.

4

CAN THE LOGIC OF INDIRECT DISCOURSE BE FORMALISED?

Recent attempts[1] to analyse or explicate assertion-statements and other kinds of indirect discourse in formal terms seem to have overlooked an even greater difficulty than the familiar problems arising about equivalence, synonymy, translatability, etc. This difficulty is created by the fact that in judicial, journalistic, or historical fact-finding, as well as in everyday conversation, we frequently cite statements about a witness's truthfulness alongside the report of his testimony as the premisses from which we argue about the facts. It is then an essential part of our argument that it should appear to jump from one level of statement to another, and perhaps back again. So that the difficulty emerges: how can the formalisation of such an argument adopt any of the usual hierarchy-principles as a guarantee against semantical antinomies?

Consider, for example,

(1) If the policeman testifies that anything, which the prisoner deposes, is false, and the prisoner deposes that something, which the policeman testifies, is true, then something, which the policeman testifies, is false and something, which the prisoner deposes, is true.

As a statement in colloquial English (1) could reasonably be called a logical truth, since the only words occurring 'essentially' in it, to use Quine's terminology,[2] are words with the analysis of which formal logicians have professionally concerned themselves, viz., 'and,' 'if,' 'then,' 'is', 'anything which,' 'something which', 'true,' 'false,' and 'that.' Anything that is substituted throughout for 'the policeman testifies' or 'the prisoner deposes,' and keeps (1) meaningful, also keeps (1) true. *I.e.*, the pattern of (1) is

(2) If A that anything, which B, is false, and B that something, which A, is true, then something, which A, is false and something, which B, is true,

and informally it is easy enough to prove that the consequent of such a conditional is deducible from its antecedent. But formalisation turns out to be much more difficult.

An informal proof might run as follows. The premises are

(3) A that anything, which B, is false

and

(4) B that something, which A, is true.

It is required to prove from these that

(5) Something, which A, is false

and

(6) Something, which B, is true.

It will first be proved from (3) and (4) that

(7) Anything, which B, is false

is false, and it will then be proved from (3) and the falsity of (7) that (5) and (6) are both true:

According to (4)

(8) Something, which A, is true

satisfies the relative clause of (7). Hence, if (7) is true, (8) must be false (by universal instantiation), and the contradictory of (8), viz.,

(9) Anything, which A, is false

must be true. Now, according to (3), (7) satisfies the relative clause of (9) ; so that, if (9) is true, (7) must be false (by universal instantiation). Hence, if (7) is true, (7) is false; and therefore (7) is false. Now, if (7) is false, its contradictory (6) is true. And, since, according to (3), 7) satisfies the relative clause of (5), it follows that, if (7) is false, (5) is true (by existential generalisation). But (7) is false. Therefore both (5) and (6) are true.

But no sentence on the pattern of (2) could even be formulated – let alone proved – in a formalised language that was constructed according to either of the two familiar proposals for analysing indirect discourse.

According to one such proposal, advocated in particular by Church, the analysans of a statement reporting someone's belief, assertion, etc. in the form of indirect discourse should be formulated in a non-extensional object language: the analysans reports a relation of belief or assertion between a person and an intensional entity like a proposition. Now, in Church's development of intensional logic intensional entities are conceived to belong in a hierarchy of successively higher orders, since the name of each such entity must have as its sense an intensional entity of higher order.[3] But it seems impossible to assign the proposition that is the sense of an analysans for (7) or the proposition that is the sense of an analysans for (9) to any order whatever in such a hierarchy. For in an analysans of (7) we should quantify over a universe which, as we are told in (4), includes the sense of (8) and therefore also presumably the sense of (9) – (8)'s contradictory; so that the sense of (7) must belong to an order higher than that to which the sense of (9)" belongs. At the same time, in an analysans of (9) we should quantify over a universe which, as we are told in (3), includes the sense of (7) ; so that the sense of (9) must belong to an order higher than that to which the sense of (7) belongs. And it is impossible for the sense of (7) to belong to an order both higher and lower than that to which the sense of (9) belongs.

According to another proposal, advocated in particular by Carnap, the analysans of a statement reporting anything in the form of indirect discourse should be formulated in a metalanguage : the analysans reports a relation of some kind (the precise nature of which is still disputed) between a person and a sentence of the object-language. On this approach the hierarchy principle adopted is the language-level policy. Each symbol or formula used or mentioned in the analysis must belong to one or more distinct languages which are so constructed, and so assigned to numbered levels, that statements about the semantics of any language assigned to level n may only be expressed in a language assigned to level $n+1$. But according to such a policy, since (4) implies that the semantics of (8) are a topic of (7), the analysans of (7) must belong to a language of higher level than the analysans of (8) and that of its contradictory (9). At the same time, since (3) implies that the semantics of (7) are a topic of (9), the analysans of (9) must belong to a language of higher level than that of (T). *I.e.*, if a normal language-level policy is to be applied in the formal analysis of statements patterned on (2), the analysans of (7) must be higher than that of (9) and the analysans of (9) must be higher than that of (7) – which is impossible. (Analogous difficulties arise if we substitute a ramified type structure for a language-level policy. For then any property designated by a function corresponding to 'A' in (2) must be both lower and higher in order than

the property designated by a function corresponding to 'B' - which is impossible. Nor would an axiom of reducibility help us at all, since it would presumably state 'To any function designating a property of any order and any type there corresponds...' (whereas we cannot regard the properties in question as belonging to any order at all).[4] There are perhaps four main ways of trying to get round this *prima facie* impossibility:

(A) We might reject any informal criterion which leads us to accept a statement like (1) as a logical truth. But this would be methodologically questionable for two reasons. First, we should not be solving any problems thereby, but simply dismissing certain problems because we cannot solve them in the same way as certain others. Secondly, the criteria, like Quine's, which lead us to accept (1) as a logical truth axe, broadly speaking, those which have always determined the subject-matter of formal logic from Aristotle's time to the present day. Someone might, of course, object that in this respect 'that' is not in the same category as the other words which occur essentially in (1) since, while they can occur in connection with any predicates whatever, 'that' can only occur in connection with predicates introducing indirect discourse, like 'testifies,' 'doubts,' etc. He might therefore maintain that (1) is not a truth of logic but merely analytic of 'that.' But such an objection would have to be backed up by the further contention that at least one of the three terms 'true,' 'false,' and 'proposition' is not to be classed along with 'and,' 'anything which,' etc. as a term of logical interest, since an analogous problem is created even if we eliminate any reference to speech, direct or indirect, from (2) by reformulating it as

(10) If 'Any Y proposition is false' is an X proposition, and 'Some X proposition is true' is a Y proposition, then some X proposition is false and some Y proposition is true.

even if this further contention seemed plausible, so that (2) and (10) were called patterns of 'analytic,' rather than of specifically logical, truth, we should still have on our hands the problem of how to formalise such demonstrably true statements.

(B) We might agree to accept (1) as a logical or, more broadly, as an analytic truth, but regard it as too trivial a truth for us to be worried by its exclusion from a formalisation of indirect discourse. But there are at least four reasons why it would be unsatisfactory to shelter under the excuse of triviality here, even if that excuse is available in other cases, like 'If this statement is true, this statement is true.' First, enquiry among those who are not professional logicians soon elicits the fact that (1) is not self evident to every educated person. If one substitutes 'denies' for 'testifies' and 'doubts' for 'deposes,' (1) becomes even

less self-evident to many people. Secondly, we can construct an infinite number of patterns of logical (or analytic) truth much more complex than (2) which for analogous reasons are equally resistant to formalisation according to any usual hierarchy principle. Thirdly, statements like (1) express logical principles on which people stand trial for perjury and on which historians sort out conflicting testimonies in their source material. If a formal logician were to reject such principles as trivial he would abdicate the importance of his own subject over a very wide field of rational discourse. Fourthly, although recently considered problems about synonymy or equivalence are no doubt among the fundamentally important ones in regard to the logic of indirect discourse, (1) is a good deal more logically complex than some of the statements that have to be considered in connection with these problems about synonymy and equivalence.

(C) If we accept that statements on the pattern of (2) deserve analysis in our formalisation of indirect discourse we might aim to avoid adopting any hierarchy-principle that would create trouble. On the intensionalist approach we should then need (*pace* Church) to assign all intensional entities to a single order of reality: on the other approach we should need (*pace* Carnap) to assign all sentences to a single level of language. But how then could we prevent the occurrence of semantical antinomies? In the one case there would be nothing to bar a sentence like 'The proposition hereby asserted is a concept of falsehood': in the other, nothing to bar 'This sentence is false.' The best we could do would be to prohibit specifically each sentence that we know to generate a semantical antinomy. But this gives us no guarantee whatever that other such antinomies might not arise to vitiate the quality of our formalised language as a deductive system – at least if that language were as rich, and its sentences as unrestricted in length, as would be required to analyse all statements containing indirect discourse.[5]

(D) We might seek to achieve such a guarantee, despite abandoning reliance on a hierarchy-principle of any kind, by restricting each formalised language we construct to the terminology and length of sentence this language required in order to afford an analysis of some given piece of informal argumentation. If we had only to operate with a formalised language of this poverty-stricken character its vocabulary would in most cases be insufficient for the formulation of any sentence generating a known semantical antinomy. Where it was sufficient for this, we should need rules specifically banning each such sentence: and with so restricted a vocabulary and length of sentence there would be a reasonable likelihood that we should spot, and thus be in a position to bar by further specific prohibitions, any sentences generating new and unfamiliar kinds

of semantical antinomy. *I.e.*, for each analysis the interpretation of the connectives, quantifiers, etc. of our underlying logistic system would be the same, but in creating a formalised language adequate to the intensionalist analysis of (1), for example, two predicate letters alone would need interpretations – as 'testified by the policeman' and 'deposed by he prisoner,' respectively. Hence, though our formalised language would then be perfectly adequate to the analysis of (1), neither 'This statement is false' nor ' "Heterological" is heterological' could be formulated in it. And when the content of our analysandum required the interpretation of one predicate letter as, say, 'heterological,' we should merely need an additional rule barring the interpretation of any formula of the logistic system as the latter sentence. This procedure does not require us to abandon the ideal of showing that a single system of logical principles underlies all rational argument : the more comprehensive formalisation normally envisaged is not the only way of achieving that ideal. For, as we pass from one analysis of this modified kind to another, the logistic system underlying our formalised language and the interpretation-rules that provide it with a logical vocabulary remain invariant - only the interpretation rules that provide it with its non-logical vocabulary need, in most cases, to vary. Indeed, this is roughly the procedure adopted in practice by most elementary text-books of formal logic, even if their authors have in mind the possibility of a more comprehensive formalisation, as is evident, for instance, from the examples and exercises in Quine's *Methods of logic* (1952). But those who wanted to accept such a solution of the difficulty would have to give up the more comprehensive idea, the idea of a single formalised language, or hierarchy of formalised languages, in which one could articulate every non-trivial logical (or analytic) truth that can be uttered in a natural language like English. The point of view they would have to accept is that a natural language is a system of human institutions which lends itself, amongst other things, to the construction of any desired number of mutually unrelated theories and arguments, and that it is these theories and arguments, with their varying structures, terminologies, and contents, not the natural languages, with their varying degrees of conceptual richness, which severally constitute the proper objects of formalisation. This solution has a number of other considerations in its favour which would require too much space to discuss here.[6] But its adoption obviously requires the renouncing of many locutions which most modern formal logicians have come to accept without question, such as 'true in English,' 'satisfies in English the sentential matrix "...",' 'valid in English,' 'the logical inconsistency of colloquial English,' etc.

Statements like (1) present so glaring an obstacle to the usual enterprise of comprehensive formalisation that one feels there must be good reasons why this obstacle has so long remained unnoticed, especially since the somewhat analogous problems created by the occurrence of impredicative definitions in classical mathematics were given comparatively early attention.[7] It seems possible that one reason has been the way in which, from *Principia mathematica* onwards, modern expounders of the Liar have tended to dissociate that paradox from indirect discourse, thus following Eubulides of Megara rather than Epimenides of Crete.[8] It has been supposed that the point of the Liar can be fully articulated by discussion of self-referential statements like 'This statement is false,' just as the point of the Grelling and Berry paradoxes can be fully articulated by discussion of self-referential predicates. On such a supposition truths like

(11) If 'Any statement by the prisoner is false' is a statement by the policeman, and 'Some statement by the policeman is true' is a statement by the prisoner, then some statement by the policeman is false and some statement by the prisoner is true

ought to have created difficulties analogous to those with which (1) confronts us. But, since people often find it difficult to voice quotation-marks without risk of being misunderstood, truths like (11) represent patterns of argument that are rather less common in ordinary speech than those represented by truths like (1) . Hence it is perhaps not so surprising that no serious notice has been taken of them. Perhaps if the Epimenides version of the Liar had received as much attention as the Eubulides one, the difficulties about logical truths like (1) might have shown earlier that the Liar indicates a problem not merely about what we should take to be proper techniques of formalisation but also – in the light of (D) above – about what we should take to be its proper objects.

Notes

1. E.g., R. Carnap,*Meaning and necessity* (1947) p. 83 ff. ; A. Church, "On Carnap's analysis of statements of assertion and belief", *Analysis* 10 (1950) p. 97 ff., and "Intensional isomorphism and identity of belief", *Philosophical studies* 5 (1954) p. 65 ff.; I. Scheffler, "On synonymy and indirect discourse", *Philosophy of science* 12 (1955) p. 39 ff.; etc.

2. *Mathematical logic* (1951) p. 2 (cf. the same author's "Truth by convention" in *Philosophical essays for A. N. Whitehead*, 1936, p. 90 ff.) If what Quine says here is to afford, informally, a necessary and sufficient criterion of logical truth under all circumstances he needs to make certain further stipulations, such as that in most cases (which need to be further specified) the same word occurs vacuously more than once only if it can be replaced by a word having precisely the same sense and reference on each occurrence in the statement. But these stipulations are not specially relevant to the purpose of the present paper and are therefore not discussed here.

3. "A formulation of the logic of sense and denotation" in *Structure, method, and meaning* (1951) ed. P. Henle *et al.* p. 3 ff.

4. In any case our axiom would have to be rather stronger than Russell's, and its justification correspondingly even more questionable (cf. F. P. Ramsey, *The foundations of mathematics*, 1931, p. 28 ff.), since, in order to formalise indirect discourse we should need (for familiar reasons) an order containing properties designated by functions that were not merely extensionally equivalent to all functions designating properties that belonged to other orders of the same type, but also synonymous with them. Moreover, it appears that even Russell's Axiom permits some of the semantical antinomies to reappear unless the names of some functions are omitted. (cf. I. M. Copi "The inconsistency or redundancy of *Principia mathematica*" in *Philosophy and phenomenological research* 11, 1950, pp. 190-9 and the review of this by A. Church in this JOURNAL, vol. 16 1951, pp. 154-5). Yet such an omission would seriously impoverish any formalised language that sets out to be as rich as the everyday natural languages in which indirect discourse is ordinarily articulated.

5. Cf. L. J. Cohen and A. C. Lloyd, "Assertion statements", *Analysis* 15 (1955) p. 68. The main argument developed in *Assertion statements* reinforces, though is less powerful than, the present paper's argument about the difficulty in principle of formalising indirect discourse. But in (D) of the present paper a way round this difficulty is discussed which was not mentioned in *Assertion statements.*

6. I allude in particular to certain problems about the interpretation of modal calculi which I hope to discuss elsewhere. The procedure discussed here is also indispensable for a third kind of proposal for formalising indirect discourse, unlike either Church's or Carnap's, which is sometimes found in the literature. In the analysans for, say, 'x asserts that p' the prefatory phrase 'asserts that' or 'x asserts that' is sometimes not represented by a predicate at all neither by an object-language function taking as one of its arguments the name of an intensional entity nor by a metalinguistic function taking as one of its arguments the name of a sentence in the object-language but by a non-extensional operator forming statements out of statements. A paper by Los on a system of this sort was reviewed in this Journal 14 (1949) pp. 64-5, and another such system is proposed by A. N. Prior in *Time and modality* (1957) p. 130-1. I have been unable to consult Los's paper and Prior's somewhat summary description of his system does not indicate explicitly how he intends to avoid the occurrence of semantical antinomies in interpreting it. But it is clear that no hierarchy principle controls the way in which Prior's operators form statements out of statements, since one category of these operators forms truth-functions of the statements they precede and even those that are prefaces to indirect discourse may precede any other such preface. Indeed, in a formalised language based on this system the analysis of (1) would be quite straightforward. The analysans is

$$CK\gamma\Pi pC\delta pNp\delta\Sigma pK\gamma ppK\Sigma pK\gamma pNp\Sigma pK\delta pp$$

where 'γ' and 'δ' are interpreted as the operators 'the policeman testifies that' and 'the prisoner deposes that,' respectively; and the formal deduction of the consequent from the antecedent of this conditional marches step by step with the informal deduction set out above. But a policy like that outlined in (D) must be pursued if '$\Sigma pK\delta pNp$' is not to generate a contradiction when 'γ' is interpreted as 'I am now asserting that' (with 'now' in a narrow sense).

7. When a set is referred to by an expression that contravenes a hierarchy principle like Russell's theory of types or Quine's stratification requirement, another mode of referring to the set may be sought that is unobjectionable on these grounds. But the referentially opaque quality of indirect discourse precludes in general any analogous method of avoiding difficulties in formalising indirect discourse according to a hierarchy principle. One may, however, observe a rough and distant analogy between the way in which (D) above seeks to avoid the semantical antinomies by specifying conditions under which a predicate letter in a syntactically well-formed formula is to be assigned a meaning and the way in which set theories in the Zermelo tradition seek to avoid the logical antinomies by setting up conditions under which a predicate is to be viewed as having an extension.

8. Cf. S. C. Kleene, *Introduction to metamathematics* (1952) p. 39.

5

SOME REMARKS ON GRICE'S VIEWS ABOUT THE LOGICAL PARTICLES OF NATURAL LANGUAGE

In the earlier part of a stimulating series of William James Lectures at Harvard in 1968 Professor H. P. Grice drew the attention of the philosophical public[1] to a most intriguing hypothesis about the familiar logical particles of natural language 'not', 'and', 'if ... then ... ', and 'either ... or ... '. I shall henceforth call this the Conversationalist Hypothesis. What it asserts is that those particles do not diverge in meaning, or linguistic function, from the formal-logical symbols, '$\sim$', '&', '$\rightarrow$', and '$\vee$' respectively, as standardly interpreted by two-valued truth-tables, and that wherever they appear to diverge from truth-functionality the appearance is due to the various standing presumptions with which natural language utterances are understood. On the whole Grice argued in favour of this hypothesis, though he confessed to having no answer to one particular objection to it. I shall argue in this paper that the objection to which Grice refers is not, *pace* Grice, a serious one, but that there are good reasons for preferring an alternative account, which I shall call the Semantical Hypothesis, to the Conversationalist Hypothesis. According to the Semantical Hypothesis many occurrences of these particles do differ in significance from their formal-logical counterparts and many do not, and both kinds of occurrence are best explained within the bounds of an adequate semantical theory for natural languages and without recourse to a theory of conversational presumptions.

Grice in fact also included such expressions as 'all', 'some', and 'the' within the scope of the hypothesis that he was supporting. But he offered no arguments about these expressions. So, like him, I shall confine my remarks on the subject to the level of propositional logic. The question

of quantification-theoretic idiom is also an interesting one, but it can be left for another occasion.

No doubt philosophers have sometimes mistakenly attributed to the meaning of a word or the analysis of a concept some feature that is more correctly regarded as a condition for the appropriateness of certain utterances involving that word or concept. Such philosophers have been struck, for example, by the oddity of discussing whether or not an action is voluntary when the action itself is a perfectly satisfactory one; and then they have mistakenly traced this oddity to the meaning of the word 'voluntary' instead of to the conditions for there being some point in remarking of a (voluntary) act that it was voluntary. No doubt there are a number of philosophical errors than can thus be corrected by paying a proper regard to the presumptions of normal conversation. But it is also possible for the pendulum to swing too far in this direction. What is better accounted a feature of linguistic meaning may sometimes be put down to conversational presumptions. This, I shall argue, is what the Conversationalist Hypothesis does in regard to the logical particles of natural language.

One can easily see why the Conversationalist Hypothesis is so tempting. If there are divergences of meaning between 'not', 'and', 'if ... then ...', and 'either ... or ...', on the one hand, and their familiar formal-logical counterparts, on the other, the task of representing the logic of natural language becomes more complicated. If this representation is to be accomplished within a formal theory one or other of two things has to be done.

On one alternative some tailor-made non-truth-functional system has to be constructed with temporal connectives, intensional conditionals, and so on. But at best such a system achieves fidelity of representation in a certain area of language only at the cost of sacrificing conceptual economy and computational facility. More commonly the system throws up its own, more subtle, divergences from natural language.

On the other alternative, the formal theory is offered as a reconstruction, rather than as a description or replication of natural language, and divergences are explained away as being unimportant for logical purposes. For example, if 'if ... then ...' is always reconstructed as the truth-functional '$\ldots \rightarrow \ldots$', most intuitively valid patterns of deductive inference that involve conditional statements, like modus ponens, can be reproduced, and no inference from true premises to false conclusions will ever be validated. But then the system also throws up inferences that seem to have no counterparts in natural language, like that from 'q' to '$p \rightarrow q$'. To avoid either alternative, and thus escape through the horns of the dilemma, one has to surrender altogether the search for a formal

representation and be content with informal descriptions of ordinary usage. But this is to sacrifice all the rigour of treatment, and opportunities for insight, that formal systematisation can provide. No wonder, then, that the Conversationalist Hypothesis should be so tempting. All these difficulties loom over us as a result of our assuming that 'not', 'and', etc. commonly differ in meaning from their familiar formal-logical counterparts. If that assumption could be safely abandoned, the classical truth-functional systems of Frege and Russell would afford an accurate and economical representation of the fundamental features of the logic of natural language.

Unfortunately, however, tempting though the Conversationalist Hypothesis may be, there are good reasons for rejecting it in favour of the Semantical Hypothesis. I shall discuss each of the four main particles in turn.

1. NOT

According to the usual two-valued truth-table definition of '$\sim$', '$\sim p$' is false whenever 'p' is true. But there are several dialects or natural languages in which 'not', or its counterparts, do not behave in this way. In several London dialects of English, for example,

(1) You won't get no beer here

is an emphatic reformulation, not a negation, of

(2) You'll get no beer here.

I.e., in these dialects (1) is true, not false, when (2) is true. Similar emphatic uses of the negative particle occur in Italian (e.g., 'Non fa niente'), in Spanish and in Homeric and classical Attic Greek.[2]

How could the Conversationalist Hypothesis be defended against such *prima facie* counter-examples? One possible move would be to say that the Hypothesis is not to be construed as making a claim about natural language in general but only about some dialects of some natural languages, e.g., Standard English. But not only is this not how Grice apparently envisaged the hypothesis that he was discussing. It also substantially weakens the claim that is made. The claim now is that some natural languages or dialects are fundamentally truth-functional, rather than that all are. Instead of purporting to express a general truth about the logic of natural language, it now does no more than describe an idiosyncrasy of vocabulary that is allegedly present in some languages and absent in others.

Another possible move would be to say that the Conversationalist Hypothesis is to be construed as making a claim not about the negative

particle 'not', and its counterparts in other languages, but rather about the phrase 'it is not the case that' and its equivalents. It may well be that those phrases do conform to the usual truth-table definition of the formal-logical constant '$\sim$'. However, this is again not how Grice himself apparently envisaged the hypothesis that he was discussing. And in any case the claim made is substantially weakened. It applies only to certain phrases in which 'not' occurs rather than to that word in general.

So perhaps a third possible defence could be offered, viz., that the apparent equivalence between (1) and (2) is not an equivalence of sentential meaning, but an equivalence in force of utterance, due to the bearing of conversational presumptions on our assertions. Now, Grice held that, cetera paribus, participants in a conversation will be expected to observe a general principle that runs roughly as follows: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk-exchange in which you are engaged." Several more specific maxims, on Grice's view, will yield results, more or less in accordance with this general principle, e.g., ' Make your contribution as informative as is required', 'Don't say what you believe to be false', 'Don't say what you lack evidence for', 'Be relevant', 'Avoid obscurity, ambiguity or unnecessary prolixity', and 'Be orderly'. It is conceivable therefore that in every case in which, in some dialect or other, someone utters a double negative as in (1) with the purport of a single negative like (2), the speaker should be construed to be speaking ironically. That is, we have to presume his obedience to the maxim 'Don't say what you believe to be false', and should therefore infer, from the obvious falsehood of what he says if taken literally, that he must be speaking ironically. There is an implicature, as Grice calls it now,[3] of irony: i.e., one has to assume irony in order to maintain the supposition that the cooperative principle and maxims are being observed. If this were the case, the negative particle in (1) would have its usual meaning, in accordance with the truth-table definition of '$\sim$', and the equivalence between (1) and (2) would not be an equivalence of sentential meaning, but the kind of equivalence that exists between an ironical utterance of 'He's a fine person to trust' and a non-ironical utterance of 'He's a bad person to trust'.

But the defence of irony will not work here. First, it doesn't fit the facts. It assumes a quite incredibly wide prevalence of ironical speech. Indeed it implies the use of irony on very many occasions on which the hearer has no reason to believe that in its literal meaning what the speaker says is false, and therefore no adequate reason to believe that he is talking ironically. Secondly, the defence of irony will be even less plausible when the emphatic double negative occurs in a subordinate clause,

as in the antecedent of a conditional or in indirect speech. Compare, e.g.,

If I won't get no beer here, I'll have a cider instead

or

They've tried to fix it so you won't get no beer here.

But, if irony is not at work here, it is very difficult indeed to see in what other way conversational presumptions could lead to the equivalence of uttering (1) with uttering (2) when 'not' is purely truth-functional. Indeed it looks as though this equivalence stems from an equivalence of sentential meaning between (1) and (2).

2. AND

Perhaps it will be said that at least Grice's own dialect is one in which 'not' always functions in accordance with the truth-table definition of '$\sim$', and never just adds emphasis to another negative. But what about 'and'? Let us look at some of the data.

Two facts seem incontestable. On the one hand, in some cases the utterance of two sentences conjoined by 'and' asserts more than just the truth of both statements. For example, there is an important difference between what is implied by an assertion, *tout court*, of the sentence

(3) A republic has been declared and the old king has died of a heart attack

and what is implied by an assertion, *tout court*, of the sentence

(4) The old king has died of a heart attack and a republic has been declared.

The order of events implied by an assertion of (3) is the converse of that implied by an assertion of (4). On the other hand, in some cases the assertion of two sentences conjoined by 'and' implies no more than the truth of both statements, as in

(5) The old king has died of a heart attack and a republic has been declared, but I don't know which of these two events preceded the other nor do I wish to suggest some connection tends to exist between two such events.

In these latter cases every bit of additional information that might have been conveyed by the utterance of 'and' is somehow cancelled or deleted. (Compare too the patent truth-functionality of 'It is the case that ..., and it is the case that ...'.)

But, though both facts seem incontestable, their interpretation is highly controversial.

According to the Conversationalist Hypothesis, which Grice favours, the meaning or linguistic function of 'and', even in (3) and (4), is just the same as that of the truth-functional constant '&', and the implication of temporal sequence that is conveyed by uttering (3) or (4) derives from a presumption that people's discourse obeys the maxim 'Be orderly'. This presumption may be rebutted, as in (5), and then no sequence or connection will be implied (or implicated, as Grice calls it). But when the presumption is not rebutted, it will always operate, as in isolated assertions of (3) or (4).

According to the Semantical Hypothesis, however, the meaning or linguistic function of 'and', as a clause-concatenating particle, is rather richer than that of the truth-functional constant '&'. In addition to expressing the conjunction of two truths it also indicates that the second truth to be mentioned is a further item of the same kind, or in the same sequence, or of a kind belonging to the same set of commonly associated kinds of item, or etc. etc., as the first truth to be mentioned. Hence an implication of temporal sequence arises in cases like (3) or (4), or of connectedness in cases like

(6) Tom has a typewriter and he types all his own letters.

But this additional feature in the meaning of 'and' and of its equivalents in other languages is subject to cancellation or deletion in certain contexts, as in (5). I.e., according to the Semantical Hypothesis what is cancelled is a feature that is one of those features which should be listed in any adequate dictionary entry for the word, whereas according to the Conversationalist Hypothesis what is cancelled is a presumption of human discourse in general – a presumption that it would be out of place to state in the description of any particular natural language.

Now how are we to choose between these two quite different theories about the correct interpretation of utterances asserting sentences like (3)? Note first that neither theory has any advantage over the other in respect of lexicographical simplicity. At times Grice seems to suggest that anyone who opposes the Conversationalist Hypothesis about 'and', 'if ... then ...', etc. must be prepared to accept at least two dictionary meanings for such logical particles - a stronger meaning that is not purely truth-functional and a weaker meaning that is. But such a suggestion would be incorrect. The Semantical Hypothesis accords just as well as the Conversationalist Hypothesis with Grice's excellent recommendation not to multiply senses beyond necessity. Both assign only one dictionary sense to 'and', as a word in the language. But the one

theory assigns a weaker sense to the word and fills out this analysis by invoking a deletable conversational implicature of orderliness, while the other theory assigns a stronger sense to the word and allows a certain feature of this sense to be deleted on occasion, in the process of composing the meaning of a compound sentence out of the meanings of its constituent words and clauses.

Moreover, while the two theories disagree about the nature of the deletion or cancellation that has to be imputed to assertions of (5), each type of deletion is certainly realized in some other, relatively uncontroversial cases. For example, if Professor X is asked his opinion of Professor Y's abilities and replies

(7) He is a very competent bicyclist,

he implicates, without its being part of his sentence's meaning, that Professor Y is not so competent in academic matters; and this implicature would be cancelled or deleted if Professor X added

(8) But I do not wish to suggest that he is not also very competent at his work.

On the other hand, the statement

That is a flower

implies, in virtue of the meaning of the word 'flower', that the object in question forms or formed a part of a plant; and this implication is cancelled or deleted if the word 'plastic' is put before 'flower'. In that respect the phrases 'stone lion', 'well-painted hand', 'sub-vocal speech', etc. are all rather similar: in each case part of the normal meaning of the noun is cancelled or deleted by the adjacent adjective.[4] Dictionaries do not need to list both a strong sense of 'flower', in which we speak of flowers as growing, and a weak sense in which we speak of 'artificial flowers', 'plastic flowers', or 'toy flowers'; and similarly the Semantical Hypothesis does not need to suppose both a strong and a weak sense for 'and'. In other words, while we can regard the Conversationalist Hypothesis as assimilating the interpretation of (5) to the paradigm of (7) + (8), we can regard the Semantical Hypothesis as assimilating it to the paradigm of

That is a plastic flower.

Is there any reason for preferring one pattern of assimilation to the other? A marginal advantage seems to be gained by the Conversationalist Hypothesis because it treats locutions like 'I do not wish to suggest that ...' as accomplishing the same type of deletion or cancellation in both

(5) and (7) + (8). However, the assumption of this argument is that similar expressions should perform similar roles and on that assumption there is an argument of about equal strength for the Semantical Hypothesis. The particle 'but' obviously has an adversative, and not merely conjunctive, function in such sentences as

> Tom has a typewriter but he prefers to write all his letters.

So, by analogy, it seems reasonable to suppose that the particle 'and' has a connexive, and not merely conjunctive, function in (6). Indeed the temptation to treat 'and' as being like 'but' in this respect (i.e., in being not merely conjunctive) becomes stronger when we consider cases like

> Tom picked up a stone, and threw it, but missed the tree, and hit a window, but the window did not break, and Tom's father was not angry.

where connection seems to alternate with opposition. It seems a little strained to suppose that in uttering such a sentence the connection conveyed by each utterance of 'and' is a consequence of conversational maxims while the opposition conveyed by each utterance of 'but' is a consequence of linguistic function. However, though it is a little strained to suppose this, it is not very difficult, just as it is also not very difficult to suppose that locutions like 'I do not wish to suggest that ...' accomplish different kinds of deletion in (5) and (7) + (8), respectively. So this line of reasoning does not seem to be conclusive in either direction.

But there is at least one type of 'and'-occurrence that the Conversationalist Hypothesis can hardly be stretched to fit at all. This is where 'and' occurs in the antecedent of a conditional and the truth of the statement made by asserting the conditional depends on the precise sequence of events suggested by this occurrence of 'and'. Consider, for example,

(9) If the old king has died of a heart attack and a republic has been declared, then Tom will be quite content

where Tom might not be at all content if a republic had been declared first and then the old king died of a heart attack.[5] Of course, it would be open to Grice to extend or adapt his theory of implicatures in such a way that anyone who asserts a conditional like (9) would normally implicate that the truth of the conditional as a whole is partly dependent on a condition of temporal sequence that is conveyed, in virtue of conversational presumptions, by the utterance of the antecedent. But this would hardly be consistent with the claim made by the Conversationalist Hypothesis, which Grice supports, that 'if ... then ...' has just the same purely truth-functional meaning as the standard formal-logical

connective '$\rightarrow$'. For, if the truth of (9) as a whole is just a function of the truth-values of its antecedent and consequent, and if 'and' is purely truth-functional, so that the truth-value of (9)'s antecedent is just a function of the truth-values of its constituent conjuncts, it follows that the truth of (9) cannot depend at all on any condition of temporal sequence that may be conveyed, in virtue of conversational presumptions, by the utterance of the antecedent. That is, the truth-functionality of 'and' in cases like (9) could only be maintained at the cost of sacrificing the truth-functionality of 'if ... then ...' (and accepting a good deal of extra complexity into the theory of implicatures). Hence, so far as the purpose of maintaining the truth-functionality of 'and' is to defend the Conversationalist Hypothesis, there seems to be no point in maintaining it: the Conversationalist Hypothesis seems incapable of being defended in this way.

Perhaps it will be claimed instead that utterance of (9) implicates the order of the conjoined clauses in the antecedent to be somehow essential to the relevance of one part of the indirect evidence for (9). (By 'indirect evidence', for a conditional, is meant evidence other than the truth-values of the conditional's antecedent and consequent.) But that would amount to claiming that an assertion of (9) implicates the existence of more indirect evidence for its truth than is necessary on a purely truth-functional account of the meaning of 'and'. It would be as if to say that the truth-functionality of 'if ... then ...' can only be maintained at the cost of sacrificing the truth-functionality of 'and'. The Conversationalist Hypothesis has again to be rejected.

3. IF ... THEN

Grice suggested, in his Lectures, that there are some cases in which 'if ... then ...' quite obviously does not diverge in linguistic function from the standard formal-logical connective '$\rightarrow$', other cases in which it may seem to diverge but can be shown not to, and yet other cases in which it seems to diverge and he, Grice, does not see how to show that it does not. Let us consider these cases in turn, as data to be taken into account in making a rational choice between the Conversationalist and Semantical Hypotheses. It should be borne in mind that on Grice's view, i.e., according to the Conversationalist Hypothesis, the assertion of an 'if ... then ...' sentence, while truth-functional in linguistic meaning, commonly carries with it a (cancellable) implicature that there is indirect, i.e., non-truth-functional, evidence for its truth. I suppose that this is because of the maxims requiring a speaker to have evidence and forbidding him unnecessary prolixity: e.g., if his only evidence for 'if

p, then q' were the fact stated by 'q', he would do better to say just 'q'. On the other hand, according to the Semantical Hypothesis, a dictionary entry for the particles 'if ... then ...' should state that they indicate a connection between antecedent and consequent as well as performing the purely truth-functional role of ruling out the conjunction of the antecedent's truth with the consequent's falsehood, though the Hypothesis is quite consistent with finding certain occurrences of these particles where the context is such as to delete the connexive aspect of their dictionary meaning, analogously to 'and' in (5).

First, then, let us consider the cases in which 'if ... then ...', on Grice's view, quite obviously does not diverge from '$\rightarrow$'. These are, on his view, the cases in which the alleged implicature of there being indirect evidence for the conditional is cancelled by the linguistic or environmental context of utterance.

Suppose, for example, two partners announced at the beginning of a game of bridge a special Five No Trumps Convention, whereby a call of 'Five no trumps' signified the statement

(10) If I have the king of hearts, I also have a black king

as well as the usual undertaking to make five no trumps if there is no overbid. Grice claimed that, on such an occasion, not only would (10) be patently truth-functional, but also its utterance would not convey the usual suggestion made by a conditional statement that there is some reason for believing it other than knowledge of the truth-values of its antecedent and consequent clauses. However, the trouble with this example is that the speaker, by uttering the call 'Five no trumps', does give the other players an indirect reason to believe in the truth of (10), as a statement in which 'I' refers to himself. Even if what he said to the others was actually (10) instead of 'Five no trumps', he would still be giving them thereby an indirect reason to believe in its truth. The only kind of utterance of (10) at the bridge table that would not count as an indirect reason to believe in its truth would be if a player said (10) silently to himself. But this he would hardly be likely to do. So even if we grant to Grice that his indirect-evidence implicature is normally cancelled when sentences like (10) are at issue, that is not because the utterer of such a sentence does not convey clearly enough to his hearers the existence of indirect evidence for the truth of his assertion, but rather because he conveys this altogether too clearly: his very utterance constitutes the evidence. Nevertheless, an assertion of (10), in the specified circumstances, does seem patently truth-functional, when contrasted with an assertion of, say

If I have the king of hearts, I am lucky.

But this is not because there is no suggestion of indirect evidence, since there is such a suggestion. Rather, it is because there is no suggestion of a connection between antecedent and consequent. Hence the Semantical Hypothesis fits such cases perfectly well. We simply have to suppose that the connexive aspect of the linguistic meaning of 'if ... then ... ' is deleted or cancelled by the context of utterance. It is rather like pointing to something in the corner of an oil-painting and saying

(11) There's a hand.

The context of such an utterance of (11) would make it perfectly clear that 'hand' here did not mean flesh and blood.

Just the same is true of another example of Grice's – the kind of logical puzzle in which you are given the names of a number of persons in a room, their professions, and their current occupations, and a few clues as to how these fit together, and then you have to determine which person belongs to which profession and is currently occupied in what. For example, a sentence like

(12) If Mr. Tailor is a cobbler, Mr. Baker is currently gardening

may be given as a clue. But the very giving of it as a clue is a perfectly adequate indirect reason to believe in the truth of the statement it makes about the puzzle situation. At the same time the nature of the puzzle context is such as to delete or cancel the connexive aspect of the meaning of 'if ... then ... ', if the Semantical Hypothesis is correct.

Similarly, if we consider a case in which according to Grice the implicature of indirect evidence is explicitly cancelled by the linguistic context of utterance, we shall see that the Semantical Hypothesis again fits perfectly well. Suppose someone says

(13) I know just where Smith is and what he is doing, but all I will tell you is that if he is in the library he is working.

Grice's view seems to have been that by uttering the clause

(14) I know just where Smith is and what he is doing

the speaker cancels the usual implicature that there is indirect evidence for the truth of the conditional

(15) If Smith is in the library he is working.

But whatever happens to the implicature, according to Grice's theory, it is certainly true that by uttering (14) along with (15) a speaker would create indirect evidence for (15). And in any case, if the utterance of

(14) deletes the assertion of a connection when (15) is uttered (and I am not sure that this deletion will always occur when (13) is uttered), this deletion can be explained in the usual way by the Semantical Hypothesis.

Thus far I have been agreeing with Grice that the assertion of certain conditional sentences may sometimes be wholly truth-functional in meaning, while arguing that the Semantical Hypothesis can cover all such cases. In this respect I class two other examples of Grice's along with (10), (12), and (13), viz.,

> If England win the first Test, they will win the series, you mark my words

and

> Perhaps if he comes he will be in a good mood.

Grice also cited two further sentences where he thought the implicature of indirect evidence would be either absent or cancelled and the assertion of the conditional would be purely truth-functional in character, viz.,

(16) See that, if he comes, he gets his money

and

(17) If he was surprised, he didn't show it.

But (16) and (17) are even worse evidence for the Conversationalist Hypothesis than (10), (12), or (13), because neither would commonly involve a purely truth-functional conditional. If the conditional in (16) were purely truth-functional, the instruction expressed by (16) as a whole would be fulfilled by seeing that the man got his money even if he did not come. But the employer who uttered (16), and thus told his cashier to pay the man if he came to work, might be inclined to dispute the view that the cashier had done his duty by paying the man even though he had not come to work. As for (17), what one has to ask, if (17) is alleged to be truth-functional, is why someone who accepts (17) because its consequent is true may nevertheless be reluctant to assert

> If he was not surprised, he didn't show surprise

of which the consequent would also be true. The reason, surely, is that 'if' in (17) has the sense of 'even if', not of 'if . . . then . . . ' as in (10), (12), or (13). You can put 'even if' in place of 'if' in (17) without changing the sense, but if you do this in (10), (12), or (13) you change the sense. 'Even if . . . ' does not normally function to state a condition that, if it holds, generates a certain consequence, as in the ordinary conditional,

but rather to state a condition that does (did, will) not prevent a certain consequence. Hence it looks as though (17) is not strictly relevant to the present discussion.

Let us now turn to cases where, on Grice's view, the implicature of indirect evidence is not cancelled. In these cases, according to Grice, the conditional 'if ... then ...' is purely truth-functional, and the suggestion of non-truth-functional reasons for accepting the conditional is carried not by the meaning but by the implicature that conversational presumptions generate. A typical case, I suppose, would be

(18) If the government falls, there will be rioting in the streets.

Now, in the case of simple, straightforward conditionals like (18) there seems nothing to choose between the Conversationalist Hypothesis, favoured by Grice, and the Semantical Hypothesis which claims the standard linguistic meaning of 'if' to be stronger than a purely truth-functional account can represent.

But consider what happens when (18) goes into the antecedent of another conditional like

(19) If it is the case both that if the government falls there will be rioting in the streets, and also that the government will not fall, then the shopkeepers will be glad.

According to the Conversationalist Hypothesis, (19) would normally carry the implicature that there is indirect evidence of its antecedent's not being true without its consequent's also being true. But for the antecedent of (19) to be true, we must have both (18) true and the antecedent of (18) false. Hence, if (18) is to be understood as a truth-functional conditional, in accordance with the Conversationalist Hypothesis, its inclusion in the antecedent of (19) would be quite inessential. To discover the truth-value of the antecedent of (19) all we need to know is the truth-value of the antecedent of (18). For, if the antecedent of (18) is false, the antecedent of (19) must be true, in virtue of the law '$\sim p \rightarrow ((p \rightarrow q) \& \sim p))$'; and, if the antecedent of (18) is true, the antecedent of (19) must be false, in virtue of the law '$(p \rightarrow \sim ((p \rightarrow q) \& \sim p)$'. It follows that indirect evidence for the truth of (19) would be quite sufficient if it related the fate of the government: to the feelings of the shopkeepers without having any bearing whatever on the causes and effects of rioting in the streets. But this is quite counter-intuitive since it assumes the truth of the consequent to be dependent on only one condition–the fate of the government–whereas in asserting (19) one would assert it to be dependent on two mutually independent conditions. Therefore, if we feel that the evidence for (19) must tell us something

about the question of rioting, we cannot accept that the occurrence of 'if ... then ...' in (18) is purely truth-functional in meaning.

Perhaps it will be objected that because of the conversational maxim forbidding unnecessary prolixity there is a presumption that in uttering (19) no inessential clauses have been uttered. Accordingly, it will be said, there is an implicature that the content of (18), as well is its truth-value, is essential to the truth of (19). But that would make (19) not wholly truth-functional – contrary to what the Conversationalist Hypothesis asserts. So perhaps the objector will instead claim an assertion of (19) to implicate that the content of (18) is somehow essential to the relevance of part of the indirect evidence for (19). But that would amount to claiming that an assertion of (19) implicates the existence of more evidence for its truth than is necessary on a purely truth-functional account of (19)'s meaning. It would be as if to say that the truth-functionality of 'if ... then ...' in its occurrence as the main logical particle of (19) can only be maintained at the cost of sacrificing its truth-functionality in respect of its occurrence within one of the conjuncts in the antecedent of (19). For sentences like (19), as for those like (9), it looks as though the Conversationalist Hypothesis is indefensible. Only some vastly complicated, and correspondingly implausible, addition to the theory of implicatures could possibly save it.

Finally we come to the cases in which 'if ... then ...' seems to diverge in meaning from the ordinary formal-logical '$\rightarrow$' and in which Grice sees no way of using his theory of conversational presumptions to show that this divergence is illusory. Consider, for example,

(20) It's not the case that, if the government falls, then there will be rioting in the streets.

On Grice's view, if assertion of (18) normally carries an implicature of indirect evidence for the truth of a truth-functional conditional, (20) must normally carry a denial of that implicature. But someone might object that when an utterance is not absurd, if taken literally, a denial of it is standardly a denial of its literal reading, not of its implicature. For example, if you say ironically

(21) He is a splendid fellow

and I reply

(22) He is not a splendid fellow

I must be saying, directly and feebly, just what you have, ironically, implicated. And to this objection Grice, in his Lectures, sees no answer.

But, compared with some of the other - above mentioned - objections to the Conversationalist Hypothesis, this one seems rather weak. The

fact that although there are some cases like (21) and (22) in which denial of an utterance is not a denial of its implicature there are also other, non-absurd utterances in which it is. For example, if I say

(23) *Someone* has not sent in a correct return

I implicate that I am unable or unwilling to tell the full story. But if I deny (23) by saying

Everyone has sent in a correct return

I also deny the implicature of (23). It looks as though Grice has over-generalized from the rather special case of ironical implicatures as in (21). Certainly, if we construe one of his conversational maxims as prescribing 'Don't say what you lack *adequate* evidence for', anyone who says p implicates that he has adequate evidence for p and anyone who denies p denies this implicature also. So, *pace* Grice, the objection under consideration looks as though it can be quite easily answered: whether or not the denial of a non-absurd assertion is also a denial of its implicature must depend on the nature of the assertion and the nature of the implicature.

Nevertheless there is a very obvious way in which (20) runs counter to the Conversationalist Hypothesis. That hypothesis holds that 'not' and 'if ... then ...' are truth-functional in meaning, though assertions of 'if ... then ...' sentences often carry with them implicatures of indirect evidence. But, if the meaning of (20) is such as to deny the truth of a truth-functional, it must logically imply both that the government will fall and that there will be no rioting; and as this implication is quite unacceptable (20) cannot be the denial of a truth-functional conditional. Grice's Lectures offered no viable rejoinder to this type of objection. It is no use rejoining that (20) should be construed as propounding a counter-conditional to (18), viz.,

(24) If the government falls, there will not be rioting in the streets,

since (24) is not at all synonymous with (20). Nor is it any use rejoining that (20) has the effect of a refusal to assert (18), or of a denial of the implicature of (18). For, if by 'effect' here is meant 'implicature' or 'illocutionary force', the rejoinder does not meet the objection, which is about the meaning of (20). And, if by 'effect' here is meant 'meaning', then the Conversationalist Hypothesis is being sacrificed in respect of the meaning of 'not' in order to try and save it in respect of the meaning of 'if ... then ...'.

4. EITHER ... OR ...

Analogous moves may be made against the Conversationalist Hypothesis in regard to 'either ... or ... ' sentences, and I shall not go into much detail here. According to the Hypothesis assertion of such a disjunctive sentence normally carries with it an implicature that there is indirect evidence for the disjunction, though the implicature may be explicitly cancelled in certain cases as in Grice's example

(25) The prize is either in the garden or in the attic, but I'm not going to tell you which,

where the speaker suggests that he knows the truth of the disjunction because he knows the truth of one of its disjuncts. But such cases are rather like (20) and (12). The speaker's utterance of (25) is an event that constitutes indirect evidence for the truth of the disjunction. Here too the speaker would not normally utter the corresponding sentence to himself, viz.

> The prize is either in the garden or in the attic, but I'm not going to tell them which,

because from his point of view there is only direct (truth-functional), and no indirect, evidence for the disjunction. So even if we grant to Grice that his indirect-evidence implicature is normally cancelled when sentences like (25) - or (10) - are at issue, that is not because the utterer of such a sentence does not convey clearly enough to his hearers the existence of such a sentence does not convey clearly enough to his hearers the existence of indirect evidence for the truth of his assertion, but rather because he conveys this altogether too clearly: his very utterance constitutes the evidence. Hence, though an assertion of (25), as of (10), would normally seem truth-functional, this cannot be because there is no suggestion of indirect evidence, since there is such a suggestion. Rather, it must be because there is no suggestion of any underlying fact or principle that limits the alternatives. Now, according to the Semantical Hypothesis, though such a suggestion is part of the dictionary-meaning of 'either ... or ... ',[6] it is deleted or cancelled in certain contexts, just as the prefixing of 'plastic' to 'flower' deletes the suggestion of forming part of a plant. So the Semantical Hypothesis can accommodate sentences like (25) just as well as it can accommodate sentences like (10) or (12).

Moreover it is easy to construct 'either ... or ... ' examples that present the same kind of difficulty to the Conversationalist Hypothesis as does (19). Consider

(26) If the prize is either in the garden or in the attic, and in fact it is in the attic, the gardener will be glad.

On the truth-functional account the disjunction is quite inessential to the antecedent of (26). Indirect evidence for the truth of (26) would be quite sufficient if it related the prize's being in the attic to the feelings of the gardener without having any bearing whatever on the question where the prize would be if it were not in the attic. But this is quite counter-intuitive, because it assumes the truth of (26)'s consequent to be dependent on only one condition - the prize's being in the attic - whereas in asserting (26) one would assert it to be dependent on two mutually independent conditions. If therefore we feel that the evidence for (26) must tell us something about where the prize would be if it were not in the attic, we cannot accept that the disjunction in (26) is purely truth-functional in meaning.

Someone may object that (9), (19) and (26) are all special cases because they mention mental attitudes in their consequents. But it does not seem difficult to construct appropriate examples of other kinds such as, for (9),

> If a shilling is pushed into the slot and the red button is depressed, a bar of chocolate will fall into the tray

or, for (19),

> If it is the case both that if the wind blows the cradle will fall, and also that the wind will not blow, then the cradle is insecurely fastened but the wind will nevertheless not interfere with it

or, for (26),

> If it is the case both that he will either jump off the ice or fall through it and also that he will in fact jump off it, then he will have saved his own life.

I conclude that there are several rather stronger objections to the Conversationalist Hypothesis than the single objection that Grice himself was worried about in his Lectures. In general that hypothesis breathes the same spirit as earlier attempts to resolve philosophical puzzles about truth by reference to speech-acts of endorsing or conceding, or to resolve puzzles about goodness by reference to speech-acts of commending or commanding.[7] Such theories gain what support they seem to have from the consideration of relatively simple examples. Their weakness becomes apparent when more complex sentences are examined – especially sentences where the locution in question occurs within the antecedent of a conditional.

Notes

1. I do not know whether Grice still holds the views that he expressed in these lectures, and I certainly do not wish to imply that he does. In any case I hope that he will take it as a tribute to the interestingness of his ideas that I have not delayed my own discussion of them until the oral publicity of the William James Lectures has been compounded by a printed version. I have taken great care not to misrepresent the content of Grice's lectures as they have been reported to me. But if there are any details on which I have erred I do not think that they affect the main point I am trying to make.

2. E.g. Euripides *Andromache*, line 986: "Ouk estin ouden kreisson oikeiou philou", literally 'There is not nothing better than ...', meaning, in Standard English, 'There is nothing better than ...'.

3. Grice's concept of implicature was explained by him in 'The Causal Theory of Perception', *Proc. Aristotelian Soc.*, Suppl. vol. 35 (1961), 121-152, but there he still used the word 'implication' for it.

4. Compare perhaps also 'He unintentionally insulted her' and 'The girl on the dust-cover is naked'. On the relation of such cancellations or deletions to other processes of semantic composition, cf. L. Jonathan Cohen and Avishai Margalit, 'The Role of Inductive Reasoning in the Interpretation of Metaphor', *Synthese* 21 (1970), 469 ff. It emerges that an order of relative importance has in any case to be supposed for the set of distinctive features that characterise a particular meaning. So the less important features will normally be the ones that are exposed to cancellation or deletion in literal usage: e.g. the prefix 'plastic' deletes the notion of growth implicit in the meaning of 'flower' but not the feature of outward appearance. Where one of the more important semantic features is deleted, we tend to regard the usage as metaphorical -e.g., in 'A child is a fragile flower'. Correspondingly it is the truth-functional core of meaning in the logical particles that is undeletable: we never ascribe such words a metaphorical usage. Of course, it might be objected that every word for a representable object or event x should be assigned a second dictionary meaning as 'representation of x'. But anyone who was prepared to multiply dictionary meanings on this scale would hardly be entitled to jib at assigning a weaker (purely truth-functional) meaning to 'and', 'if ... then ...', 'either ... or ...', etc., in addition to a stronger (connexive) meaning. Also, presumably, the decision between listing one meaning or two in the dictionary entry for 'and' must be matched by a corresponding decision, in discourse analysis, with regard to the semantics of sentence concatenation.

5. I have borrowed this example, in a modified form, from my *The Diversity of Meaning*, 2nd ed., 1966, p. 271.

6. I am trying to suggest here no more than the general nature of the non-truth-functional element in the meaning of 'either ... or ...' I am not offering, in this article, an exact lexicographical characterisation of 'either ... or ...', any more than of 'and' or of 'if ... then ...'. But I certainly do not wish to claim, as is sometimes claimed, that a dictionary entry for the locution 'either ... or ...' should mention as a feature of its meaning that it indicates the speaker's ignorance of which alternative is true. Where an utterance of the locution does indicate this, the indication seems to belong rather to what Grice calls the implicature than to the meaning.

7. Cf. my 'Speech-Acts', in T. Sebeok (ed.), *Current Trends in Linguistics*, vol. XII (forthcoming).

6

CAN THE CONVERSATIONALIST HYPOTHESIS BE DEFENDED?

In the earlier part of his William James Lectures at Harvard in 1968 H. P. Grice examined the hypothesis that certain standing presumptions of normal conversation can account for all the apparent ways in which the meanings of the familiar logical particles of natural language ('not', 'and', 'if ... then ...' and 'either ... or ...') diverge from the standard two-valued interpretations of the corresponding formal-logical symbols ('$\sim$', '&', '$\rightarrow$' and '$\vee$'). Subsequently I argued that this hypothesis, which I called the Conversationalist Hypothesis, was substantially more implausible than Grice himself admitted.[1] R. C. S. Walker has now tried[2] to defend the Conversationalist Hypothesis against my criticisms, and in the present paper I aim to show that Walker's attempted defence of that hypothesis is open to several serious objections. I shall begin by recapitulating briefly (I) the content of Grice's Conversationalist Hypothesis, (II) my main line of argument against it, and (III) Walker's rejoinder to this argument. I shall then (IV) describe several reasons for dissatisfaction with Walker's rejoinder.

I

According to Grice there is a standing presumption that participants in a conversation will observe a general principle that runs roughly as follows: 'Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk-exchange in which you are engaged.' Several more specific maxims, on Grice's view, will yield results more or less in accordance with this general principle, e.g., 'Make your contribution as informative as is required', 'Don't say what you believe to be false', 'Don't say what you

lack evidence for', 'Be relevant', 'Avoid obscurity, ambiguity or unnecessary prolixity', and 'Be orderly.' The presumption of conformity to such maxims brings it about that we normally convey more than we say when we say that p – specifically, that our saying that p has an 'implicature', as Grice calls it. For example, when a man says that p he normally implicates that he has evidence that p, that the statement that p is relevant to the topic of conversation, that he does not believe it to be false that p, and so on. Hence on Grice's view the apparent difference between

(1) A republic has been declared and the old king has died of a heart attack

and

(2) The old king has died of a heart attack and a republic has been declared

might be accounted for not by supposing that the conventional, linguistic sense of 'and' involves an element of sequence, but by bearing in mind the implicature of order which an utterance of (1) or (2) would have: 'and' can then be purely truth-functional in its sense, like '&'. Similarly, when a man says that

(3) If the government falls, there will be rioting in the streets,

his saying this normally has the implicature - generated by the combined effect of various maxims - that there is indirect evidence for (3): i.e. that there is evidence other than facts which are just about the truth-values of the antecedent and consequent of (3). Hence, though the assertion of (3) would normally suggest a connexion between its antecedent and its consequent, this suggestion is to be accounted for not by the meaning of 'if' but by the implicature of the assertion. The truth-functionality of 'if' is thus preserved.

II

Grice admitted to being disturbed by one possible objection to the Conversationalist Hypothesis about the logical particles of natural language. But I argued in my earlier paper that, while this objection was in fact invalid, there were much more serious objections which Grice had not noticed. These objections arise when sentences like (1), (2) or (3) occur in the antecedents of certain kinds of conditional sentence. Consider, for example,

(4) If a republic has been declared and the old king has died of a heart attack, then the king's supporters will revolt

and

(5) If the old king has died of a heart attack and a republic has been declared, then the king's supporters will revolt.

Informed diplomats might well think (4) to be true and (5) false. Yet if 'and' and 'if ... then ...' are both fully truth-functional (4) and (5) should have the same truth-value.

Again, consider

(6) If it is the case both that if the government falls there will be rioting in the streets, and also that the government will not fall, then the shopkeepers will dance for joy.

On a straightforwardly truth-functional analysis (3) could be omitted from the antecedent of (6) because (6) must have the same truth-value as

(7) If the government will not fall, then the shopkeepers will dance for joy.

But informed diplomats might think (6) true and (7) false.

There are various ways of trying to save the truth-functionality of some occurrences of logical particles in propositions like (4), (5) and (7) at the cost of sacrificing that of others. But the conclusion of my earlier paper was that the Conversationalist Hypothesis as a whole could be saved only by 'some vastly complicated, and correspondingly implausible, addition to the theory of implicatures'.

III

Walker claims that these arguments do not show the Conversationalist Hypothesis to be unsatisfactory. His key move is 'to distinguish the truth-value of an utterance from the truth-value it would have if one were to consider only its propositional content'.[3] Indeed, he says, if the conversationalist were not allowed to draw this distinction, 'his Hypothesis would be a non-starter', since he would have to admit

(8) If I stop writing this essay now, Amin will conquer Kenya

as 'true when it is palpably false (unless we jettison intuition and import a technical notion of literal truth)'. What the Conversationalist Hypothesis requires, according to Walker,[4] is that the truth-values of complex utterances '*would* be completely determined by the truth-values of the component sentences if conversational conditions could be abstracted from'. But often, according to Walker[5] the context shows that this abstraction cannot be made: we then have to be 'prepared to take into

account what a speaker conveys conversationally, as well as the sense and reference of his remarks, in determining the truth-value of what he says'. So by calling certain logical particles truth-functional, writes Walker,[6]

> the conversationalist does not wish to claim that what we should normally, in everyday talk, think of as the truth-value of the complex utterance is determined truth-functionally by the truth-values we should ordinarily and naturally assign to the constituent sentences. If he were to claim this the Conversationalist Hypothesis would be so wildly implausible as not to require the lengthy treatment given to it by Cohen; for it is quite obviously *false* in any normal acceptation, that if I stop writing this essay now the Queen will abdicate, though on a narrowly truth-functional account it would be true.

Accordingly, Walker's defence of the Conversationalist Hypothesis in regard to a complex conditional like (4), (5) or (6) is to argue that when determining the truth-value of the antecedent as a whole we must take into account not just its propositional content but also its implicature. When we do this it is easy to see how (4) can have a different truth-value from (5), and (6) from (7). In a statement like (4) or (5), 'the truth-value of the conditional is now being said to depend not just on the truth-values of antecedent and consequent, but on something else: the order of the events mentioned in the antecedent'.[7] Similarly in (6), according to Walker,[8] though the sense of the conditional contained in its antecedent is purely truth-functional, what this contained conditional conveys is 'not truth-functional at all', and when we consider its truth-value we have to bear in mind not only its sense and reference but also the conversational implicatures of its utterance.

IV

There was a great deal more in Grice's lectures than I have suggested, and there was also considerably more in my own paper and in Walker's. In Sections I-III above I have just abstracted certain relatively isolable strands of disagreement, and I propose to continue in the same vein now by arguing that Walker's defence of the Conversationalist Hypothesis is unsatisfactory. I shall not say anything further here about the Semantical Hypothesis which my earlier paper set up as an alternative that was preferable to the Conversationalist Hypothesis. Nor shall I discuss Walker's claim[9] that the notion of a conversational implicature 'looks likely to be of enormous value in the resolution of philosophical disputes', since to discuss this claim one would need to broach a rather large question about how much the philosophy of x is, or should be, concerned with local patterns of conversation about x. Philosophers who do not treat their subject matter as primarily oral are perhaps less likely

to indulge in the confusions that a theory of conversational implicatures might resolve; and it is not immediately clear which of Grice's maxims are maxims to which any communication whatever must normally be presumed to conform, and which maxims are relative to a particular culture or to a particular form of communication. Also there are strong reasons[10] for supposing that the elucidation of metaphor should be carried on within a semantics for natural language, and not, as Walker suggests,[11] by a theory of implicatures.

(1) The first point to note about Walker's attempt to defend the Conversationalist Hypothesis is that it misses its mark. It is not a defence of the hypothesis that Grice advanced and I criticised, but of a rather different hypothesis. Grice implied throughout his William James Lectures that the truth-values on which logical particles operate are always the truth-values of component propositions, and are never determined by the implicatures of uttering these propositions.

Grice distinguished, in his second Lecture, what a man 'says' from what he implicates by saying it, so that 'what is said may be true - what is implicated may be false', as he remarked at the end of the lecture. Correspondingly where something false is conveyed by his saying something true, Grice did not want to say that his 'utterance of this sentence would be, strictly speaking, false'. Correspondingly, in his fifth Lecture, he discusses a case where the remark

(9) X or Y will be elected

is countered by the reply

(10) That's not so: X or Y or Z will be elected.

But he construes (10) to be rejecting (9) 'not as false but as unassertable': (9) is rejected, not because both its disjuncts are false, but because its utterance falsely implicates the existence of adequate indirect grounds for the disjunction. Yet Walker would certainly regard this as a case where (9) was being treated as false, and not merely unassertable, in virtue of its false implicature.

For he claims that 'our inclination to let truth depend on what is implicated varies with the likelihood that the speaker will want to be committed to it'.[12] Indeed, where the affirmation of 'If p, then q' would carry an implicature, Grice (in his fifth Lecture) confessed himself unwilling to accept the thesis that the denial of the conditional is the denial of that implicature. Yet he took the plausibility of this thesis to be a serious objection to his Conversationalist Hypothesis. So he must have firmly intended his hypothesis to imply that logical particles operate on constituent propositions in abstraction from the implicatures of their

utterance. This is confirmed by Grice's express attitude (in his fourth Lecture) to the possibility that in certain circumstances the applicability of 'true' and 'false' might be 'situation-relative', as he called it. If the application of these terms ever depended on the particular circumstances in which a statement was made, he claimed, then that would be a good reason for interpreting '*T*' and '*F*' in the truth-table definition of a logical particle as 'factually satisfactory' and 'factually unsatisfactory', respectively, rather than as 'true' and 'false'. Even at the cost of sacrificing the normal interpretation of the truth-tables, the definitions of the logical particles had to preserve their independence from any consideration of implicatures. And Grice made the same point quite generally towards the end of his second Lecture. 'Since the calculation of the presence of a conversational implicature presupposes an initial knowledge of the conventional force of the expression the utterance of which carries the implicature, a conversational implicatum will be a condition which is not included in the original specification of the expression's conventional force'.

Thus the version of the Conversationalist Hypothesis which Walker finds 'so wildly implausible as not to require the lengthy treatment given to it by Cohen' is in fact Grice's own version: I was at least in good company when I thought it deserved a lengthy treatment. Moreover my arguments against Grice's Conversationalist Hypothesis cannot be invalidated by the fact they do not apply to Walker's Conversationalist Hypothesis, since the two Hypotheses differ in the crucial respect upon which Walker's counter-argument relies.

(2) Nevertheless, if Walker's Conversationalist Hypothesis were in fact tenable, this itself would be quite important. Grice's Hypothesis would have been saved, as it were, by modification. So Walker's Hypothesis seems to deserve investigation in its own right. But investigation of it reveals a number of difficulties.

The most important of these difficulties is that the truth-values of logically compound utterances are not always functions of the truth-values of their component propositional utterances. The truth-value of a logically compound utterance, like the utterance of (1), (2), (3), (4), (5), (6) or (7), is determined, according to Walker's Conversationalist Hypothesis, by the truth-values of its largest component propositional utterances; but the truth-values of these latter utterances, where they in turn are compound, are determined not only by the truth-values of their own component propositional utterances but also, where contextually appropriate, by the truth-values of their implicatures. In this way the truth-value of (6)'s utterance, for instance, comes to be determined in part by the implicatures of uttering its antecedent, and accordingly the truth-value of this antecedent's utterance is not wholly determined by the truth-values of its ultimate propositional-utterance components. Thus

Walker's Conversationalist Hypothesis sacrifices the objective that was central to Grice's Conversationalist Hypothesis, viz. the maintenance of a thoroughgoing truth-functionality for logically compound assertions of the kind that we have been considering. Specifically, on Grice's Conversationalist Hypothesis, which attaches truth-values to what is said, truth-functionally equivalent components may be substituted for one another in compound assertions like (4), (5) or (6) *salva veritate*: on Walker's Conversationalist Hypothesis, which attaches truth-values primarily to the saying of things, this substitutivity is no longer guaranteed. Walker himself does not seem to be aware of this fact since he writes[13]

> There is something very attractive about the Leibnizian idea that the propositional content, the sense and reference, of what we say can be expressed in a formalised language for which all the legitimate inferential moves are precisely determined by syntactical means; someone who considers himself to have independent grounds for holding that this formalised language must be truth-functional will naturally adopt the Conversationalist Hypothesis.

What Walker's Conversionalist Hypothesis does maintain is a truth-functional meaning for the logical particles - in the sense that the truth-value of a logically compound assertion is determined by the truth-value or truth-values of the utterance or utterances upon which the compounding particle immediately operates. But that is a far cry from the maintenance of a truth-functional analysis for the assertion as a whole.

(3) It might still be argued however that on this score Walker's Conversationalist Hypothesis is no worse than what I called, in my own paper, the Semantical Hypothesis. The latter is the view that the logical particles of natural language have stronger meanings than a purely truth-functional analysis can deploy, though in certain contexts the extra, non-truth-functional element may be cancelled, much as in

That is a plastic flower

the term 'flower' loses the implication that its denotatum forms or formed part of a plant. But, though Walker's Conversationalist Hypothesis is certainly not inferior to the Semantical Hypothesis in respect of its non-truth-functionality, it does generate certain paradoxes and complexities which the Semantical Hypothesis avoids.

One obvious paradox lies in Walker's assignment of falsehood to any utterance of a true proposition which commits the speaker to a false implicature. For example, a man might say truly that the car involved in an accident was his neighbour's and yet lack evidence for this. But when we knew the falsehood of the implicature to which he had committed himself we should not normally describe his utterance as 'false'. 'True, though misleading' would be a much apter description. Worse still, what happens when the implicatures to which a speaker is apparently

committed do not all have the same truth-value? For example, the man who says that the car involved in an accident was his neighbour's might lack evidence for it and yet not believe it false. According to Grice's conversational maxims such a man's utterance has both a false and a true implicature. So on Walker's Conversationalist Hypothesis the man's utterance is both true and false at the same time, despite the fact that all the man actually said was that the car involved in the accident was his neighbour's.

(4) Again, consider the two assertions

(11) The *New York Times* correspondent said that (6) is true

and

(12) The *Washington Post* correspondent said that (7) is true

when 'said' is used in Grice's sense, whereby what a man 'says' is to be distinguished from what a man implicates by saying it. It scarcely seems bizarre or idiosyncratic to feel that one could, in certain circumstances, comment truly on this

(13) What the *New York Times* correspondent said is true and what the *Washington Post* correspondent said is false

Now Grice's Conversationalist Hypothesis does not allow the possibility that (13) could ever be true, and that is a mark against it. But Walker's Conversationalist Hypothesis can only allow the possibility of truth to (13) by refusing to accept that 'said' in (11), (12) and (13) has the restricted sense assigned to it. Walker's Conversationalist Hypothesis has to reinterpret this 'said' as being an equivalent of 'uttered', 'asserted', or some such expression that will include any contextually appropriate implicatures as well as propositional content. Yet the restricted sense of 'said' is hardly an odd one. Indeed it is not easy to expound the theory of conversational implicatures itself without making use of this restricted sense of 'say', or some such verb, which was the interpretation originally intended for 'said' in (11), (12) and (13).

(5) Finally, there is a serious risk of circularity in the exposition of Walker's Conversationalist Hypothesis. The risk is that understanding the implicatures of a logically compound utterance may be made to depend on understanding the sense of the sentence uttered, but because of the presence of a logical particle in the sentence, understanding of the sentence's sense may depend on understanding the implicatures of its utterance. In order to avoid such circularity the theory of implicatures would have to be introduced in a considerably more complicated way than Grice's Conversationalist Hypothesis requires. It would need

to be specified first how implicatures are generated by the utterance of a propositional clause in which no logical particle occurs, whether this clause be the sole linguistic component of an utterance or only one of several such components. Next, it would need to be specified how the truth-value of the utterance of such a particle-free clause relates both to its propositional content and to its implicatures. Next, the sense of a logically compound clause in which only one logical particle occurs would have to be related to the truth-value (or truth-values) of the utterance (or utterances) of its component clause (or clauses). Next, it would have to be specified how implicatures are generated by the utterance of compound clauses of this type, and how the truth-value of such an utterance is determined. And so on, in accordance with the increasing number of logical particles occurring in the utterance. Such vast complications, even if recursively specifiable, make the Conversationalist Hypothesis lose all attractiveness. If that Hypothesis cannot meet the difficulties mentioned in my previous paper without being modified along the lines Walker proposes; if it cannot then maintain its thorough-going truth-functionalism; if it cannot compensate for its non-truth-functionalism by avoiding paradoxes in judgements about the truth-value, consistency and interpretation of natural language utterances; and if it cannot compensate for this paradoxicality by its simplicity as a theoretical regimentation; then there is only one conclusion that can reasonably be drawn – the Conversationalist Hypothesis has to be discarded.

Notes

1. L. Jonathan Cohen, 'Some Remarks on Grice's Views about the Logical Particles of Natural Language', in Y. Bar-Hillel (ed.), *Pragmatics of Natural Languages*, 1971, pp. 50-68. Grice's second Lecture has now been published separately under the title of 'Logic and Conversation', in The *Logic of Grammar* (ed, by D. Davidson and G. Harman), 1975, pp. 64-75.

2. Ralph C. S. Walker, 'Conversational Implicature', in S. Blackburn (ed.), *Meaning, Reference and Necessity*, 1975 pp. 133-181.

3. *Op. cit.*, p. 153.

4. *Op. cit.*, p. 139.

5. *Op. cit.*, p. 140.

6. *Op. cit.*, p. 139.

7. *Op. cit.*, p. 138.

8. *Op. cit.*, pp. 143/4.

9. *Op. cit.*, p. 134.

10. Cf. L. Jonathan Cohen and Avishai Margalit 'The Role of Inductive Reasoning in the Interpretation of Metaphor', *Synthese* 21 (1970) 469-487, reprinted in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, 1972, pp. 722-740.

11. R. C. S. Walker, *op. cit.*, pp. 158/9.

12. *Op. cit.*, p. 140.

13. *Op. cit.*, p.. 178.

7

HOW IS CONCEPTUAL INNOVATION POSSIBLE?

1.

No one nowadays would deny the importance of conceptual innovation in the growth of scientific knowledge. But how is it possible? And by this I do not mean: what kinds of social, economic, or mental developments are causally responsible for promoting it? That is a question for historians, sociologists and psychologists of science to answer. Instead I shall concern myself with a more philosophical issue, namely: how can the possibility of conceptual innovation be compatible with the way in which we reason about language, meaning and understanding – i.e., what adjustments in, or constraints on, the framework of such reasoning are forced on us by acceptance of this possibility? In particular does it fit in with the project of reconstructing scientific reasoning in artificial languages like Leibniz proposed in the 17th century or Carnap in the 20th? Admittedly, to pose the question in those terms is to assume the legitimacy of viewing the issue from a linguistic point of view. And, if modern psychologists are right to attribute certain elementary concepts to human neonates, to chimpanzees, to domesticatable animals, etc., then language is not necessary for thought and conceptual innovation need not always involve semantic innovation. But it looks as though pre-linguistic or non-linguistic concepts are far too crude and unrefined to enter influentially into the growth of *scientific* knowledge. So we are presumably safe to treat the issue about conceptual innovation as one about word-meanings: how are new word-meanings possible?

Of course, we must be careful not to trivialise the problem. There is no doubt a sense in which every new noun-phrase or verb-phrase that

is spoken in an intelligible utterance – i.e., every phrase that has never been uttered before – expresses a new concept. Abbreviate such a phrase by a single word and you most probably have not only a new concept, in this sense, but also the *definition* of a new word-meaning. Moreover, if we accept – with current grammatical fashion – that a sentence can be of any finite length, then the procedure just described will allow an infinite number of new word-meanings. So there is no mystery at all, it might be said, about how new word-meanings are possible: they are definable in terms of old ones.

How surprising then that almost all the first occurrences of new word-meanings that are cited in the big *Oxford English Dictionary* and in other historically oriented dictionaries did not take place in definitions at all! Apart from a few specially coined words like 'gas', which are duly assigned their authors' initial definitions, and a few technical usages with deliberately stipulated meanings, the development of a new meaning is normally recorded in terms of a sentence in which the new meaning just fits the context markedly better than the old one. So it looks as though there has to be some other way in which new meanings are possible. And this cannot be by the gradual development of a habit or practice, whereby first one member of a speech-community and then others gradually come to share some appropriate tacit convention of the kind that David Lewis, for example, describes.[1] The trouble is that the new meaning is regarded by historical lexicographers as being already there in the very beginning of such a development. Its first occurrence is dated back to this beginning, so the *existence* of the new meaning cannot depend on the maturity of the linguistic practice. If the new word-meaning did not spread to other members of the speech-community, it would be relatively unimportant and might not deserve recording at all. But the one-time *existence* of the new meaning would not be retrospectively annulled thereby, and we should all the more stand in need of an explanation for its possibility.

In fact we have something of a paradox here. On the one hand, we want to be able to distinguish between the actual meaning of a word in a particular utterance – i.e., the meaning that the language requires it to have there – and the meaning that a speaker intends it to have there, in order to be able to recognise malapropisms, slips of the tongue etc., for what they are: for in any such case of lexical error the actual meaning is different from the intended meaning. And in order to maintain this distinction we need to suppose that somehow the actual meaning is socially sanctioned while the intended meaning is determined by the speaker on his own. On the other hand, it seems clear that on the occasion on which a new word-meaning makes its first appearance in the

language it cannot then represent a new social convention. Yet it seems already to belong to the language as if it did: it seems already to be an actual, and not merely an intended, meaning.

2.

To explain how this can be we need to differentiate between insulationism and interactionism as two different accounts of how the meaning of a sentence may depend on the meanings of its component words.

According to the insulationist account the meaning of any one word that occurs in a particular sentence is insulated against interference from the meaning of any other word in the same sentence. On this view the composition of a sentence resembles the construction of a wall from bricks of different shapes. The result depends on the properties of the parts and the pattern of their combination. But just as each brick has exactly the same shape in every wall or part of a wall to which it is moved, so too each standard sense of a word or phrase is exactly the same in every sentence or part of a sentence in which it occurs. We may sometimes need to look at neighbouring words in order to discover the sense in which a word is functioning in the sentence in question, as we might infer the concavity of one brick from the convexity of its neighbor. But even then the meanings that we discover are not made what they are by one another, any more than the presence of a convex brick actually alters the shape of its neighbour. Rather, the words in the sentence have been given these meanings by diachronic facts of etymology. In the case of an artificial language these meanings have been stipulated or assigned by its inventor, and in the case of a natural language they have been created by a dateable change or innovation of fashion in speakers' intentions or conventions.

Interactionism makes the contradictory assertion: in some sentences in some languages the meaning of a word in a sentence may be determined in part by the word's verbal context in that sentence, though the extent and nature of this determination shows a wide range of variation. On this view the composition of a sentence is more like the construction of a wall from sand-bags of different kinds. Though the size, structure, texture and contents of a sand-bag restrict the range of shapes it can take on, the actual shape it adopts in a particular situation depends to a greater or lesser extent on the shapes adopted by other sand-bags in the wall, and the same sand-bag might take on a somewhat different shape in another wall or in a different position in the same wall. By exploiting local context in this way a language can be much more prolific of semantic variety than insulationism can give it credit for being. Moreover these sense-refinements or sense-modifications are gen-

erated by verbal interaction within a particular synchronic state of a language. Many of them, especially metaphors, are not common enough to be recorded in dictionaries or assigned dates and places of origin. Nor are they conceived as independently existing, externally determined constituents of the meanings of the sentences in which they occur, but as arising only within particular sentences through local semantical interactions between the sentence's component words. However, such a sense-modification is sometimes sufficiently useful and occurs sufficiently frequently in actual speech that it comes to be treated as belonging to the relevant word quite independently of context. It is thought of as an alteration in the range of alternative meanings which that word has, and as being available for use in ways that an insulationist account would fit quite adequately. For example, that is how live metaphor passes into dead metaphor.

3.

In general, though with some significant exceptions, artificial languages satisfy an insulationist account whereas natural languages require an interactionist one.

The reason why most artificial languages are insulationist in their mode of semantic composition is that they are deliberately designed both to have meanings in their tokens that are quite independent of these tokens' non-linguistic contexts and also to avoid equivocations of any kind. This objective was explicitly propounded by most seventeenth-century writers on the subject of a *characteristica universalis.* Dalgarno, Wilkins, Leibniz and the others thought of themselves as seeking a new notation for scientifically worthwhile concepts, not just a new way to cipher existing words, and this new notation was to be intelligible in any human community at any time.[2] In that way they hoped that scientists would be able to avoid the confusions that were generated by the vagueness and ambiguity of existing terminology. And avoidance of all risk of equivocation becomes indispensable where, as in the Frege-Russell style, an artificial language is designed to articulate logical structure. If the burden of discerning logical validity is to be transferred from intuition to calculation, and executed by effectively checkable operations on perceivable notational shapes, then the whole procedure is simplified if in any one such exercise each pattern of shapes must be supposed to stand in for just a single pattern of word-meanings. Otherwise fallacies may be more difficult to detect. Moreover an interactionist language cannot easily avoid the possibility of equivocation if we consider only the speech-forms uttered and pay no attention to circumstances of utterance. If a word can be disambiguated, when it occurs, by its type of linguistic

context, C_1, C_2, ...or C_n, then it may sometimes be ambiguous, viz., if it can occur in some type of context other than C_1, C_2, ...or C_n. Accordingly artificial languages are normally insulationist in their mode of semantic composition in order to satisfy the purposes for which they are constructed. Some computer-programming languages are exceptions to this rule, it may be held, but this can be put down to the need for user-friendly economies in keyboard size. (Consider the role of '=' in Basic, for example.) Esperanto, Interlingua, Interglossa, etc., are also exceptions to the rule because they have been designed for everyday use, not for the rational reconstruction of scientific theories or the formalisation of proofs and arguments. In everyday speech the non-linguistic circumstances of utterance can serve to disambiguate the utterance of a sentence in which a word is not disambiguated by its linguistic context. But formal or theoretical language-systems aim at a level of precision that requires them to be independent of such environmental considerations. Those who construct these analytical languages may well seek to *represent* thereby various forms of dependence on environmental considerations (e.g., by postulating different kinds of possible worlds in which certain sentences may be asserted). But they do not thereby *replicate* such dependence: an appropriate distance must be preserved between analysans and analysandum. Certainly Tarski's method of assigning meanings to the well-formed formulas of an artificial language, for example, is rigorously insulationist. The satisfaction-conditions that are associated with occurrences of its primitive predicate-letters in atomic sentential functions are satisfaction-conditions that remain associated with the same predicate-letters in any compound formula, however long or complex. And, *mutatis mutandis*, we can say the same of Montague semantics.

Hence Montague was wrong to claim that there is no important theoretical difference between formal and natural languages.[3] Formal languages are strongly motivated towards insulationalism. But natural languages are free to use interactionist methods in order to make immense gains in their ratio of expressive potential to size of vocabulary. If by their patterns of sentential composition words can impose further subdivisions of sense on one another, in addition to the primary divisions that are recorded in dictionaries, or can impose extensions of sense beyond any yet recorded, far fewer words have to be coined and kept in mind for any given set of expressive tasks, and innumerable new tasks can be performed without any need to coin new words. If our own natural language were not already interactionist, it would pay us to make it so.

4.

I am not speaking here of homonyms, like '(enclosure)pen' and '(writing)pen', or '(legal)case' and '(suit)case'. Such homonymy, when it occurs, is a pathological phenomenon in a language, due to an accidental convergence of etymologies. Nor am I speaking of the relatively accidental variations of meaning that are not immediately intelligible in a uniform and systematic way. We get used to specialisations, like 'undertaker' in the sense of 'mortician', or exemplifications, like 'glass' in the sense of 'drinking-glass', generalisations, like 'ring', i.e., 'ring-shape', or associations, like 'bead' for the objects associated with prayer. But these meanings are all taught separately to children, or learned by them as they grow up. Perhaps they all arise out of older meanings through well-recognised diachronic patterns of innovation; but they have to be learned separately because historical contingencies have also played a part. If rosaries had not been in common use, the word for prayer would not have acquired the meaning of 'bead'. Moreover, the meanings attainable by one or other such pattern are not normally supposed to be all already present potentially in any synchronic state of the language. For example, though polythene bags exemplify polythene just as much as glass tumblers exemplify glass, we are inclined to regard

She put her shopping into a polythene,

as deviant, unlike

She poured some water into a glass.

Consider instead the use of a word to denote a two-dimensional or three-dimensional representation of the object or situation or substance that it is also used to denote, as in

(1) Four stone lions occupy the corners of Trafalgar Square.

Whether this usage exists in all natural languages, I do not know. But it is familiar in all modern European languages, and its existence in Ancient Greek was remarked by both Plato and Aristotle.[4] And the usage does not have to be learned separately for each noun denoting a species of animal, plant, artefact, event, etc. There may well be at least one such English noun that has never yet been uttered in a representational sense. But we do not have to await an utterance of this kind before being willing to accept that a representational sense already exists for it in the language.

Another similar variation of sense is that generated by the difference between occurrent and dispositional exercise of an ability, as with

'driver', 'leader', 'orator', etc., and another by the difference between act and object, as with 'perception', 'digging', 'utterance', 'regulation', etc. Of course, very many of these variations are listed in sufficiently large dictionaries, and earliest occurrences of them are recorded. But this is not done comprehensively. Thus the big Oxford dictionary[5] lists 'lion' and 'tiger' in the representational sense, and not 'rabbit' or 'zebra'. But we have to suppose that the representational meaning is present potentially in all such words, and is just waiting to be brought out by the context in a sentence in which such a word occurs as, for example, the local meaning of 'lions' is determined by 'stone' in (1). Nor is there any obvious limit to the fineness with which the meanings of certain words may be subdivided. Consider 'drop', for example, in the sentences

(2) Most students here drop geography in their final year

(where 'drop' means 'drop studying'),

(3) Most students here drop geography lectures in their final year

(where 'drop' means 'drop attending'),

(4) Most students here drop geography lectures reading assignments in their final year

(where 'drop' means 'drop executing')

(5) Most students here drop geography lectures reading assignments library-fees in their final year

(where 'drop' means 'drop paying'), and so on indefinitely. If we accept that a sentence can be as long as we please, then there seems no predictable end to the variety of expressions that we can put meaningfully after 'drop', so as to impose a series of different meanings on the latter word. And the same is true of many other verbs, like 'make', 'collect', 'establish', etc.

Note too that the variations in the denotation of 'drop' in (2)-(5) cannot be put down to vagueness or generality rather than to ambiguity or polysemy. They cannot be regarded as different species of a common genus rather than as stemming from different meanings of the same word. This is evident from such sentences as

(6) John has dropped his geography lectures and Peter has done so also.

Here 'dropped' may imply 'dropped giving' in both cases, or it may imply 'dropped attending' in both cases. But it cannot, except as a

rather poor joke, be used to denote one kind of thing in John's case and another in Peter's, though this would be quite acceptable in

(7) John has lost a pet and Peter has done so also,

where John may have lost a cat, say, and Peter a dog. Even if a sentence is vague in the way that (7) is, we can still understand it well enough to investigate whether it is true. But (6) could not be sufficiently understood without disambiguation. Similarly the attributive use of adjectives and adverbs exploits an indefinitely large range of potential meaning, in the sense that, 'good' may be said to have different, though analogous, meanings in

(8) He has a good car,

and

(9) He has a good house,

because in (8) 'good' means something like 'conforms to the standards for commending cars' and in (9) something like 'conforms to the standards for commending houses'. A good guard-dog may be a bad gun-dog, and a bad philosopher may be a good driver. Nor is this use of adjectives and adverbs confined to familiar terms of evaluation or measurement, like 'good' or 'large'. The possibility of attributive usage extends quite widely through at least the English vocabulary (though all, or almost all, words that can be used attributively can also be used non-attributively). A person who is called a 'systematic philosopher' need not be systematic in anything but his philosophy, a 'learned historian' need not be learned in anything but history, a 'comfortable river cruiser' need not be comfortable on the high seas, a substance that is a 'yellow stainer' need not be yellow itself, and an eighteenth century merchant who was a 'triangular trader' (i.e., traded around the Britain-West Africa-West Indies triangle) was certainly never triangular himself. Of course, you can say that all these usages originate from taking the agent suffix '-er' or '-an' to apply to a whole phrase, such as 'systematic philosophy', 'learned history', 'comfortable river cruise', etc. instead of just to a single word. But then it is still the semantic interaction of the adjective with the noun or noun-phrase that forces this shift of application, which we do not find in 'unemployed philosopher', 'hunchbacked historian', 'diesel-driven river-cruiser', etc.

Moreover, while all these usages are examples of interaction in the generation of literal meaning, it is obvious that semantical interaction can also be attributed a role in the generation of metaphorical meaning - a form of meaning which has so often been vital both for literary creativity and for the progress of scientific theory. If one is an insulationist

in one's conception of natural language, there is no way to explicate metaphor plausibly within the framework of semantics, since according to insulationism every word has just to be assigned its own fixed meaning or meanings. One has then to think of metaphor as a field of study for pragmatics: it arises in human speech from the relation between a sentence uttered and certain expectations imposed by the circumstances of its utterance. But for interactionists metaphor is just a rather extreme form of that kind of semantic refining or shaping which appears in so many different forms also in what is regarded as non-metaphorical usage. Consider such a series of phrases as 'drops a marker-flare in the sea', 'drops a sign to the captain', 'drops a hint to a friend', 'drops a clanger in the interview', and 'drops a bomb on the proposal'. On a normal reading of these phrases their metaphoricalness seems a matter of degree, which increases with the growing remoteness of the underlying analogy. So an interactionist point of view allows the indisputably literal and the indisputably metaphorical to be just opposite poles of a continuum within the indefinitely creative potential of natural-language meaning. But an insulationist point of view breaks this continuity and extrudes metaphor into the same range of pragmatic concerns as irony, allegory, litotes, etc., where a speaker is supposed to say one thing and to intend to convey something else.

5.

Of course, to claim that natural languages are interactionist is not to claim that every pattern of sentence composition exhibits the influence of every word's meaning on every other. If insulationism did not fit, more or less, in very many sentences, no one would ever have supposed, as many logicians have supposed, that it fits in all cases. For example, Davidson's project for a truth-theoretical semantics for natural language has a specious plausibility of this kind. To build its T-sentences on the basis of satisfaction-conditions for its primitive predicates such a semantics has to operate within an insulationist framework: the meanings of these predicates must remain the same whatever the verbal context in which they occur. And there are innumerable sentences of natural language in which it seems quite plausible to assume that the whole pattern of composition is insulationist. What blocks attempts to supply a truth-theoretical semantics for natural language is not that all patterns of natural-language sentence-composition are interactionist but just that some are.[6]

Once the difference between insulationism and interactionism has been recognised it becomes clear that we cannot construct a semantics for any natural language along the same lines as a semantics for a formal

system of any currently familiar kind. Projects like Davidson's or Montague's cannot succeed. And one can see the temptation for philosophers of language to relapse into a later-Wittgensteinian emphasis on anomalism. Of course, we can still use interpreted formal systems as logically articulate reconstructions of arguments or theories that have been propounded initially in natural language. That is to say, we can *translate* utterances in our everyday interactionist language into sentences of a suitable insulationist one. But we must not treat such translatability as a licence to identify the semantics of our everyday language with the semantics of the interpreted formal system into which we translate it. For example, Frege imputed one interactionist feature to natural language, when he claimed that, where an expression occurs in indirect discourse, it refers to the sense that it has when it occurs in direct discourse, and does not refer to what it refers to in direct discourse: correspondingly an expression's sense also is different in indirect discourse. Obviously, if this is what is happening, the sense of an expression is being changed by its linguistic context in accordance with a systematic rule that is inherent in any synchronic language-state.

But, when Church formalised Frege's logic of indirect discourse,[7] he introduced symbols with subscripts denoting types of sense, so as to be able to have a sequence of distinct symbols in his formal language corresponding to the sequence of types of occurrence of each natural language predicate viz., occurrences in its direct sense and reference, in its indirect sense (i.e., referring to its direct sense), in its indirect indirect sense (i.e., referring to its indirect sense), and so on. Church's formal language could thus be thoroughly insulationist, even though it was designed for the specific purpose of formalising an interactionist feature of Frege's semantics for natural language.

Frege's interactionism was confined to the semantics of indirect discourse, where in any case alternative analyses are available, and had nothing to say about issues like those raised by (2)-(5) above, where no insulationist analysis seems possible. But in the eighteenth and nineteenth century psychology of association a point was sometimes made that came close to anticipating the distinction between insulationist and interactionist semantics. Thus Hartley claimed that, when a large number of simple ideas coalesce into a complex one, this complex idea may appear to bear no relation at all to its component parts, just as in a highly compounded medicine 'the several tastes and flavours of the separate ingredients are lost and overpowered by the complex one of the whole mass' and the whole mass 'has a taste and flavour of its own, which appears to be simple and original, and like that of a natural body'.[8] And Mill too held that 'the laws of the phenomena of mind

are sometimes analogous to mechanical, but sometimes also to chemical laws'.[9] A complex idea that is formed by the quasi-chemical blending together of several simpler ones, should be said, on Mill's view, to result from, or be generated by, the simple ideas and not to consist of them. One example he gave was the idea of Extension (or occupation of space) in which, he thought, we cannot 'discover those elementary ideas of resistance derived from our muscular frame in which it has been conclusively shown that the idea originates'.[10] What Hartley and Mill were arguing was that when a set of simple ideas coalesces into a complex one the resultant complex may not always entail each of the original simples. And this thesis has a certain parallelism to the interactionist claim that the meaning of a phrase compounded from, say, two words W_1 and W_2 may be such that the concatenation W_1W_2 does not entail W_1 because the meaning of W_1 may be affected by the presence or absence of W_2 in the same sentence.

But, of course, Hartley and Mill viewed the composition of a complex idea as a causal process, whether quasi-chemical or quasi-mechanical, while the composition of a sentence's meaning within a semantics for natural language is likely to be viewed nowadays rather as a timeless, logical operation.

6.

The thesis that natural language requires an interactionist semantics may encounter the objection that all the facts alleged to favour it can also be explained on insulationist assumptions. 'The interactionist', it will be said, 'has to treat some word-contexts as being *evidence* that the word in question has been used in a particular pre-existing sense and other word-contexts as being *influences* that impose some refinement or modification on a pre-existing sense according to certain presumably uniform principles. Homonyms like "case" or specialisations like "glass" are offered as examples of the former kind of word, whereas words like "good" or "drop" are offered as examples of the latter. But can this distinction be maintained? Why should we not suppose that the sense that the occurence of a particular word has in a particular sentence is always pre-existent, as it were, and that the word's context always operates as evidence, leaving the word ambiguous where the evidence is insufficient?'

The interactionist must certainly grant that the border line between the two types of meaning-determination is a fluid and flexible one. For example, the fact that some dictionaries list some object-names or animal-names in a representational sense suggests that it would be possible to list representational senses for all object-names, animal-names, plant-

names, etc., whether or not these had ever been actually used. Of course, one would need a recursive specification of these representational senses in the case of each word, because the word 'face', say, can mean not only a representation of a face but also a representation of a representation of a face, and so on, as becomes intuitively apparent immediately you consider what you could say about a picture of a picture gallery in which one of the pictures is also a picture of a picture gallery. But it would certainly be possible to treat the meaning-generative principle involved here as belonging not, in a single, unified form, to the theory of sentence composition but rather, in thousands of different, particularised forms, to the details of dictionary construction.

The real difficulty for the insulationist arises in the case of words like 'drop' where an indefinite variety of senses seems capable of generation, in accordance with linguistic context. No recursive specification of these senses is possible, because no ordering relation is intrinsic to them. Of course, one could generalise about such senses by saying that for any activity A, if doing A is X-ing A, then to drop A is to stop X-ing A. But anyone who agreed that such a generalisation was a necessary component of the dictionary-entry for 'drop' would be admitting thereby that 'drop' not only has to be defined contextually, but also has an indefinite number of different definitions that vary with the context. Certainly one would not be able to give any definite list of meanings for the word that could be incorporated into a dictionary entry. And it is because this kind of list is not available in such cases that it seems reasonable to suppose that the various meanings are generated by the process of sentence-composition rather than that they somehow pre-exist this process and are utilised as its raw materials. The insulationist story loses its intuitive plausibility if there is no reasonable pr pect of ever detailing all the specific word-meanings which should be included in the dictionary that is to list the raw materials for sentence composifion.

Moreover, just as it is possible to shift the borderline between the two kinds of meaning-determination so as to make the insulationist story more extensively applicable, so too one could shift the borderline in the other direction. That is to say, one could treat any lexical innovations that are shown by their linguistic context to conform to a recognisable pattern (e.g., specialisation, exemplification, or generalisation) as instances of some appropriate rule whereby meanings are generated in the process of sentence-composition.

In short, while homonyms like 'pen' and 'case' are definitely on the side of the borderline where insulationism holds sway, and indefinitely pliable verbs like 'drop' and 'make' are definitely on the interactionist side, there is a lot of polysemy in natural language that can apparently be treated

in either way because it is just not known whether all the word-meanings concerned have actually already occurred in human utterances. In such ambivalent cases considerations of theoretical simplicity are the only factors that can determine the issue. And a compositional semantics is undoubtedly simpler if a few lexical principles can be given that control all interactions between word-meanings than if the number of different meanings to be listed in the dictionary entry for each of a large number of words has to be very substantially increased. For example, there are (at least) two possible ways to deal with representational meanings. The simpler way is to specify a general rule allowing any word that names x in some contexts to name a representation of x where the context, as in (1), requires it. The more cumbrous way is to make the dubious assumption that up to some specified level of representation of representations the relevant meanings have actually already occurred in human utterances and then to list each of these separately for each of the thousands of words concerned.

7.

It has to be emphasised, of course, that the details of verbal interaction in natural language are as yet little understood[11] and that very many problems remain as yet unresolved. For example, when one word dominates another, as 'stone' dominates 'lion' in (1), or 'geography' dominates 'drop' in (2), what ensures that the domination proceeds in one direction rather than the other? Does the less ambiguous dominate over the more ambiguous, or topic over comment, or the relatively abstract over the relatively concrete, or the inanimate over the animate? But whatever be the correct solution of these problems, it seems highly unlikely that we shall obtain any guidance towards finding it from the ideas of Tarski, Montague, etc., about the semantics of artificial languages. Recognition of the difference between insulationist and interactionist conceptions forces us to treat the semantics of natural language as a largely autonomous discipline, rather than as a topic for Davidsonian or Montagueian theory.

If we do this, the way to deal with the problem of conceptual innovation becomes clear. We can now account for the possibility of new word-meanings that are not the outcome of definitional abbreviation or historical accident. Such a word-meaning may be supposed to have been immanent in the socially sanctioned interactional potential of the language until a sentence or sequence of sentences that generates it happened to be uttered. No listing of the language's vocabulary would have revealed any such item until then. But after the word-meaning has occurred once in an actual utterance it may well come to be repeated and

eventually merit recording as a customary usage or practice that would even be recognised from an insulationist point of view. So at a certain date scientists are recorded as talking and thinking about waves that are not in water, about heavenly bodies that are not immaterial, about particles of matter that are not visible, and so on. In other words an interactionist semantics is needed in order to explain how such word-meanings can arise, and an insulationist semantics represents how they are recorded in the history of ideas.

Let us examine how precisely this happens in a particular case. In the introduction to *The Origin of Species*, published in 1859, Charles Darwin remarks that

(10) As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be *naturally selected.*[12]

What does 'naturally selected' mean here? According to the *Oxford English Dictionary*[13] the verb 'to select' was used publicly prior to 1859 only to mean the action performed by a person or other volitional agent in choosing or picking out one entity in preference to another.'[14] Similarly Latin dictionaries[15] record no other use for this verb's etymological ancestor 'seligere'. So on an insulationist account the meaning of the phrase 'naturally selected' in the English sentence (8) ought in 1859 to have been the same as it would be to-day in the English sentence

> He is the best footballer in the college, so every time a team is formed he is naturally selected.

But of course that is not the meaning which Darwin intended, nor the meaning that, even in 1859, anyone would normally have read into (10). The clause 'and thus be naturally selected' will only be coherent with the argument that forms the earlier part of (10) if it means roughly the same as 'and thus be singled out by processes of nature'. So we need to adopt an interactionist account if we are to grant the occurrence of the phrase 'naturally selected' in (10) the meaning that would normally be read into it there, which is also the meaning that Darwin intended in his utterance of the sentence.

Indeed Darwin ridiculed[16] those who objected that in his theory the term 'selection' implied conscious choice in the animals which become modified or who urged that, as plants have no volition, natural selection is not applicable to them. He admitted that 'in the literal sense of

the word ... natural selection is a false term', but drew a parallel with chemists speaking of 'the elective affinities of the various elements' despite the fact that 'an acid cannot strictly be said to elect the base with which it in preference combines'.[17] We can surely agree with Darwin that 'everyone knows what is meant and is implied by such metaphorical expressions'[18] except that by now, like so many other metaphors, these have entered into our standard vocabulary, and are listed explicitly in dictionaries and we no longer feel them as metaphors. So, though an interactionist semantics is needed to explain how such meanings could be generated originally as metaphors, an insulationist semantics would suffice to model their current manner of employment in literal usage.

Perhaps someone will nevertheless object that such conceptual innovations are possible even in a wholly insulationist language. The new meaning for a word like 'selection' or 'elective', it will be said, can always be spelt out in other terms and introduced – for purposes of convenience – by definitional stipulation: indeed Darwin himself suggested how that could be done when he remarked

> This preservation of favourable individual differences and variations, and the destruction of those which are injurious, I have called Natural Selection.[19]

But the objection puts the cart before the horse. The fundamental problem is about how Darwin could be supposed to have arrived at his new concept, not about how he could have introduced it to others. The underlying act of innovation takes place in the thinking of the scientist long before he publishes his thoughts, and it is ultimately for this act of conceptual innovation that an interactionist language is indispensable. Moreover here, as elsewhere, it is analogy that lies at the root of metaphor. Darwin drew an analogy between variations produced by human plant- or animal-breeders and variations occurring in a state of nature.[20] Just as the former resulted from human selection, the latter could be said – in his own mind – to result from natural selection, and in this way a weakened, or generalised, concept of selection could be generated by abstracting from the familiar concept any connotation of volitional agency. The new concept would then embrace both kinds of variation. But the new concept has to be already available within the interactional potential of the language if the scientist is thus to discern the relevant similarities and unifolmities. Retrospectively it may be possible to construct a periphrasis out of the resources of an insulationist language, but until the initial metaphor has been formulated under the driving force of analogy it would not be clear what has to be thus paraphrased.

Accordingly there is an important kind of conceptual innovation which we cannot hope to express or reconstruct within the framework of an artificial language that admits only of an insulationist semantics. Indeed, if Dalgarno, Wilkins, Descartes, and Leibniz had been successful in their plans to persuade every scientist to use the same 'characteristica universalis' for all his thoughts and formulations, they would have stopped scientific theorising dead in its tracks. They would have effectively blocked off the springs of conceptual innovation that abound in natural language. And even the Carnapian programme for the logical reconstruction of the language of science, though not intended as a programme for replacing natural by artificial language in the discourse of natural scientists, is nevertheless inherently blind to the processes of conceptual innovation and its results are therefore correspondingly restricted in their scope and misleading in their effect as representations of the structure of actual scientific reasoning.[21]

Notes

1. D. Lewis: 1969, *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Massachusetts.

2. L. Jonathan Cohen: 1954, 'On the Project of a Universal Character', *Mind* lxiii, 49-63.

3. R. Montague: 1974, *Formal Philosophy*, R. H. Thomason (ed.), Yale University Press, New Haven, p. 188.

4. *Republic* 597b, and *Categories ad init.*

5. *Compact edition*, 1971.

6. This thesis is developed in greater detail in L. Jonathan Cohen: 1985, 'A Problem about Ambiguity in Truth-theoretical Semantics', *Analysis* 45, 129-134.

7. A. Church: 1951, 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H. M. Kallen, and S. K. Langer (eds.), *Structure, Method and Meaning: Essays in Honor of Henry M. Sheffer*, Liberal Arts Press, New York, pp. 3-24.

8. D. Hartley: 1810, *Observations on Man, his Frame, his Duty and his Expectations*, Wilkie and Robinson, London, First Part, pp. 76-78.

9. J. S. Mill: 1896, *A System of Logic: Ratiocinative and Inductive*, Longmans Green, London, p. 558.

10. Ibid.

11. One important study in this area is J. F. Ross: 1981, *Portraying Analogy*, Cambridge University Press, Cambridge.

12. Mentor edition, 1958, p. 29.

13. Compact edition, 1971, p. 2714.

14. Actually Darwin seems to have used the phrases 'the natural means of selection' and 'principle of selection' in the new sense in a manuscript of 1839, extracts from which were published in *Journal of the Proceedings of the Linnean Society of London*: Zoology III, 1859, pp. 46-50.

15. For example, C. T. Lewis and C. Short, *A Latin Dictionary*, Clarendon Press, Oxford, 1896, p. 1663, and *The Oxford Latin Dictionary*, Clarendon Press, Oxford, 1982, p. 1728.

16. Op.cit. p. 88

17. Ibid.

18. Ibid.

19. Ibid.

20. Op. cit. p. 29.

21. Earlier versions of this paper were presented at the May, 1984, conference on Davidson's work in New Brunswick, NJ., and at the May, 1985, meeting of the International Academy for Philosophy of Science in Fribourg, Switzerland. I am grateful to participants at those meetings, and also to Professor E. Kittay, for helpful comments.

8

SHOULD NATURAL-LANGUAGE DEFINITIONS BE INSULATED FROM, OR INTERACTIVE WITH, ONE ANOTHER IN SENTENCE COMPOSITION?

I

It is often assumed that the fundamental question about definition concerns how different kinds of definition – ostensive definition, stipulative definition, dictionary definition, etc. – should be structured. But, unless we first determine how definitions should be expected to function, we cannot hope to arrive at a satisfactory account of how they should be structured. We need to know how a mechanism is intended to behave before we can work usefully on its design. This paper will therefore be concerned with one important aspect of the prior issue. Specifically, when a particular set of lexical definitions, in a particular syntactic framework, is supposed to compose the meaning of a given sentence, should those definitions be supposed to operate independently of one another or not?

In fact insulation and interaction are two different ways in which the meaning of a word in a particular sentence of a natural or artificial language may be construed, within a compositional semantics for the language, as being related to the meanings of other words in the same sentence. I want first (in I-V) to expand and clarify the distinction between insulationism and interactionism that I have drawn elsewhere;[1] secondly (in VI) to draw attention to an apparently rather plausible objection to the thesis that at least some of the definitions relevant to a particular reading of a natural-language sentence should be construed

as operating along interactionist lines; and thirdly (in VII) to show why this objection is not as strong as it might at first sight appear.

In a wholly insulational system of sentence composition each word, phrase or clause in a sentence has what meaning it does have on a particular reading of the sentence quite independently of what any other sentence-part contributes to that reading. That is to say, it makes the same contribution – a contribution describable by the same definition – whatever transformation any other sentence-part contributes to that reading. That is to say, it makes the same contribution – a contributi0n describable by the same definition – whatever transformation any other sentence part undergoes. It is, as it were, wholly insulated against the influence of other parts. For example, in a Tarski-interpretation of first-order quantification-theory the instantiation conditions applicable to occurrences of a primitive predicate P in any one well-formed formula are the same as those applicable to occurrences of P in any other well-formed formula, whatever other predicates occur in the latter.

In any interactive system of sentence-composition, however, any word, phrase or clause in a sentence may – may but not must – have its meaning affected by the meanings of one or more other expression that occur elsewhere in the sentence: the meanings may interact with one another. So, albeit with the same underlying definition, a particular word or phrase may be able to make a different contribution to the meaning of a sentence in which it occurs if another part of the sentence is replaced by a different expression. Arguably one apparent example of this is when an equivocal word is precisified, as in the transition from

(1) That is a large one

to

(2) That is a large question

or to

(3) That is a large house.

In this way a natural-language sentence is like a situation in a game of chess, as Saussure suggested,[2] insofar as the potentially interactive meanings of a sentence's component words are like the potentially interactive roles of the pieces in a particular chess situation.

To change the simile one could say that the composition of a sentence along insulational lines resembles the construction of a wall from bricks that are available in any appropriately cataloguable variety of shapes. The result of the operation depends on the properties of the parts and the patterns of their combination. But just as each brick of a type listable

in the catalogue has exactly the same shape in very wall or part of a wall to which it is moved, so too each word is lexicographically listable and has exactly the same meaning in every sentence in which it occurs. We may sometimes want to look at neighbouring words in order to discover the sense in which a word may be function in the sentence in question, asa we might infer the concavity of one brick from the convexity of its neighbour. But even then the meanings that we discover are in no way *made*, or partially *made*, what they are by one another, any more than the presence of aa convex brick actually alters the shape of its neighbour. Rather, the words in the sentence have been given these meanings by external facts. In the case of an artificial language these meanings have been stipulated or assigned by its inventor, and in the case of a natural language they have been created by a dateable change or innovation of fashion in speaker's intentions or conventions.

But the composition of a sentence along interactional lines is more like the construction of a wall from sand-bags of different kinds. Though the ex-works size, structure, texture and contents of a sand-bag restrict the range of shapes it can take on, the actual shape that the sand-bag adopts in a particular situation may depend to a greater or lesser extent on the shapes adopted by other sand-bags in the wall. the same sand-bag may take on a somewhat different shape in another wall or in a different position in the same wall. And by exploiting contextual influences in this kind of way a language may turn out to be much more prolific of variations in word-meaning that insulationism can give it credit for being.

Note, however, that the difference between insulation and interaction here is a difference between different modes of sentence-composition not of mental processing. The distinction belongs to semantical theory not to psychology. Certainly important psychological issues arise about the impact of context on word-comprehension.[3] But these issues concern the processes by which, when a sentence is uttered, its word-tokens are identified and understood. They do not concern the genesis or constitution of the word-type meanings that would need to be defined in any fully articulated account of the sentence's composition.

Nor does the distinction between insulation and interaction belong to pragmatics. For example, Atlas has discussed the important question whether we should suppose that in interpreting an utterance we construct its meaning from the relatively unspecific sense of the sentence uttered, or alternatively we should suppose that we select the utterance's meaning from a set of linguistically given readings for the sentence.[4] But, whichever of these two suppositions is the correct one, the insulationism/interactionism issue still arises. That issue is independent of Atlas's

question because it concerns how the meanings of sentences in a language are composed, not how pieces of speech should be interpreted. No doubt the non-linguistic circumstances of an utterance often impose a specificity of sense on the utterance that is not attributable to the sentence uttered when considered on its own. But higher levels of specificity are also achieved by amending or extending the sentence uttered, and an important part of my problem is to determine how this is possible.

More precisely, the question that arises is this: what considerations favour a language's being construed as exploiting the one mode of sentence-composition rather than the other, and, in particular, how relevant is the difference between an artificial and a natural language?

II

Obviously the mere artificialness of an artificial language does not force an insulationist semantics. Consider, for example, the susceptibility of a word in Esperanto to undergo the same kind of interactions as those to which equivalent natural-language words are subject. But where an artificial language is designed to articulate logical structure, as is Whitehead and Russell's *Principia Mathematica*[5] an insulationist pattern seems highly desirable. If the burden of discerning logical validity is to be transferred from intuition to calculation, and executed by effectively checkable operations on perceivable notational shapes, then the procedure for doing this is considerably simplified if in any one such exercise each pattern of shapes, whatever its context, must be supposed to stand in for just a single pattern of word-meanings. If instead there is some polysemy and some expressions have to be made more specific, or disambiguated, by reference to their linguistic context, fallacies may be more difficult to detect because the process of specification or disambiguation would add a new level of complexity to the task.

Indeed, once the correct reading of a non-logical expression is made to rely in part on the meanings of other non-logical expressions within the same sentence or well-formed formula, it is not easy to prevent descriptions of this kind of relationship from running over into statements about the inferences to be drawn from the various types of *non*-linguistic context in which utterances of the sentence or well-formed formula may occur. If, for example, the sentence (1) could have its ambiguity determined by having 'house' in place of 'one', the ambiguity might in practice be just as well determined instead by the speaker's being understood to be looking at a house. But the discernment and classification of relevant factors in the non-linguistic setting is likely to be a relatively uncertain affair when compared with procedures that involve nothing but the transformation of notational patterns.

Correspondingly it is not surprising that, even when a logician believes that he can see an interactionist feature in the logical structure of natural language, he may well prefer an insulationist pattern for its formalisation in his artificial language. For example, so far as Frege intended the term 'sense', or '*sinn*', to cover natural-language meaning, he imputed an interactionist feature to natural language when he claimed that, where an expression occurs in indirect discourse, it refers to the sense that it has where it occurs in direct discourse.[6] In other words, on Frege's view the expression does not have the same reference in indirect discourse as it has in direct discourse, and an expression's sense is also analogously different in indirect discourse. And Frege certainly did not suppose that a complete lexicon for a natural language would list the indirect sense of a word alongside its customary sense, its indirect indirect sense alongside its indirect sense, etc., and that understanding indirect discourse requires us in each case to make appropriate selections from such lists. Rather, according to Frege, the sense of an expression is open to being changed by its being embedded in certain kinds of linguistic context, in accordance with a systematic principle that is inherent in any synchronic state of a natural language. But when Church formalised Frege's logic of indirect discourse,[7] he introduced symbols with subscripts denoting types of sense, so as to be able to have a sequence of distinct symbols in the resources of his formal language corresponding to the sequence of types of occurrence of each natural-language predicate, viz. occurrences in the predicate's customary sense and reference, in its indirect sense (i.e. referring to its customary sense), in its indirect indirect sense (i.e. referring to its indirect sense) and so on. Church's artificial language could thus be thoroughly insulational, even though it was designed for the specific purpose of formalising an interactional feature of Frege's semantics for natural language.

On the other hand there is a good reason why some or all word-meanings in natural language should possess interactional features. If a natural language can exploit such features according to learnable principles, it is in a position to make immense gains in its ratio of expressive potential to size of vocabulary. If through their patterns of sentential composition words can impose further subdivisions of sense on one another, in addition to the primary divisions that are recorded in dictionaries, or can impose extensions or restrictions or variations of sense beyond any yet recorded, far fewer words have to be coined and kept in mind for particular expressive tasks, and indeterminately many – perhaps infinitely many – new tasks can be performed without any need to coin new words. If our own natural language did not already possess interactional features, it would apparently pay us to endow it with them.

Correspondingly a compositional semantics that exploited interactional principles would need far fewer lexical definitions.

III

The disambiguation of homonyms, like 'bill' or 'bank', plays only a relatively unimportant part here. Such words have etymologies that exhibit, often over a period of centuries, an accidental, and potentially confusing, convergence of form. So any case of homonymy is normally to be viewed as a pathological development. Indeed, if a word is to be individuated by the etymology of its current form (rather than solely by that form), there are two words 'bill', not one. And the sentences

(4) The bill is large, though we shall have to pay it

and

(5) The bill is large, though not as large as a pelican's

are then not instances of a single word's having different meanings in different contexts and so are not instances of interaction.

Again a word also often takes on new meanings by tacit or express convention, without losing its old meanings, and then the interaction of other word-meanings within the same sentence may serve to resolve ambiguity or enhance specificity. But, so far as the innovation is a purely conventional one, use of the same word-form for the new sense is determined by principles of diachronic rather of synchronic semantics. A new word-form could instead be coined for the purpose and no disambiguation or specification by interaction is then needed. We could speak more often about the philosophy of 'empirism', for example, instead of about that of 'empiricism'.

Where instead interaction shows itself to have a systematic function in natural language is in dealing with principled polysemy. Where this exists available meanings proliferate in certain ways without any need for conventional recognition or datable innovation. Consider, for example, the use of a word to denote a two-dimensional or three-dimensional representation of the object, organism, situation or substance that it is used in its fundamental, default sense to denote, as in

(6) Four stone lions occupy the corners of Trafalgar Square.

Whether or not this usage exists in all natural languages, it is certainly familiar in all modern European ones, and its existence in Ancient Greek was remarked by both Plato and Aristotle.[8] Or consider the use of a word to denote the audible representation of a sound, or an artefact that reproduces the sound, as in

(7) The sixth symphony is at the bottom of the pile.

And a language-speaker does not have to learn these usages separately, as if they were lexicographically listable cases of homonymy or of conventional innovation.[9] He does not have to *learn* representational senses for each noun denoting a species of animal, plant, artefact, event, sound, etc., so as to be able to refer to stone tigers as well as to stone lions, to toy soldiers as well as to toy sailors, to mock weddings as well as to mock executions, to plastic leaves as well as to plastic flowers, and so on. There may well be at least one such English noun that has never yet been uttered in a representational sense. But we do not have to await an utterance of this kind before being willing to accept that a representational sense is already available for it in the language.

Perhaps someone may be tempted to object that in the utterance of sentences like (6) the speaker merely conforms to some pragmatic rule whereby expressions meaning one thing are used in a kind of shorthand speech-act to denote something rather different. According to this objection the word 'lion', without changing its meaning, becomes an instrument for talking about statues of lions, and so on. But that kind of interpretation of the situation is available only at too high a price, because it would force us to accept that (6) is true if and only if the stone figures at the corners of Trafalgar Square are real lions.

Moreover the usage extends further in at least two directions. First, it is available for suitable noun-phrases or verb-phrases, as well as for suitable nouns or verbs, as in

(8) There is a sea-battle on one wall and a land-battle on another.

Secondly, it is available for representations of representations, as becomes evident immediately you consider what you could say about a picture of a picture-gallery in which one of the pictures is also a picture of a picture-gallery. So, if in regard to the description of representations there is no limit to the length of a suitable descriptive phrase, or no limit to the possibilities of recursion, there are potentially an infinite number of additional word-meanings available here.

IV

Many other types of principled polysemy are to be found in at least some natural languages. For example, there is the ability of a word denoting a person who has a certain type of skill to denote instead a person who is employing that skill (e.g. 'driver', 'critic', 'guide', etc.). And in many sentence-types contextual interaction serves to determine which is presumptively denoted, as in the difference between

(9) Drivers are easily trained

and

(10) Drivers will pick up their passengers at the door.

Similarly words that are able to denote a certain type of thing are also able to denote an instance of that type, as in the difference between

(11) The swan is a beautiful species of bird

and

(12) The swan is swimming in the pond.

But there are also some areas of linguistic usage in which the choice between an interactionist and an insulationist semantics may seem more debatable. Consider, for example, such phrases as 'a good gardener', 'a good story', 'a good deed', or 'a good meal'. On an interactionist account we might say that in each of these phrases the word 'good', in addition to having its general evaluation sense, also have a distinctive descriptive sense. By interaction with the meaning of 'gardener' the word 'good' comes to signify – in the phrase 'a good gardener' – such features as 'knowledgeable about how to grow plants', etc. And this kind of interactionist analysis was not uncommon forty or fifty years ago. We find it, for example, in analytical treatises of that period on moral philosophy, such as those by Stevenson or Hare.[10] It implies that an evaluative word like 'good' has indefinitely many descriptive meanings available, since so many different criteria of 'goodness', etc. may be applicable. In a particular sentence the operative criterion, if not left indeterminate, is selected by context. Alternatively, however, an insulationist analysis of such phrases may be proposed. On this analysis the meaning of a term like 'good' does not change from context to context. Instead the reader or hearer of the sentence, say,

(13) George is a good gardener

is left to draw an appropriate inference about George on the basis of his or her commonsense knowledge of the features characterising a good gardener. On such a view phrases of the form 'a good ...' might be said to mean just 'a person or thing having the characteristics appropriate for commending a ...', and to lack any more specific, or descriptive, meaning.

Nor is this choice between an interactionist and an insulationist account available only for evaluative terms like 'good', 'fine', 'beautiful', etc. It is also available for other terms that are applied in accordance

with variable standards, such as 'flat', 'high', 'smooth', 'thick', etc. The difference between the flatness of a flat country and the flatness of a flat lawn may be regarded as a difference in the meaning of 'flat' if we adopt an interactionist account and as a difference in the criteria appropriate for judging flatness if we adopt an insulationist one.

An analogous choice is available for the treatment of certain verbs. Consider, for example, the difference between

(14) Most students here drop their geography lectures in their final year

and

(15) Most teachers here drop their geography lectures in their pre-retirement year.

On an interactionist account 'drop' in (14) means 'drop attending' and 'drop' in (15) means 'drop giving'. But on an insulationist analysis 'drop' means the same in (14) and (15). So on this analysis (14) is glossed by

(16) In relation to geography lectures most students in their final year stop doing what ever is standardly appropriate for students to do in relation to those lectures,

while (15) is glossed by

(17) In relation to geography lectures most teachers in their pre-retirement year stop doing whatever is standardly appropriate for teachers to do in relation to those lectures.

And in each case the reader or hearer, with his commonsense general knowledge about what is thus appropriate, can then infer more specifically what it is that most students - or teachers - stop doing in their last year.

To many the interactionist construal of sentences like (13) or (14) may seem less intuitive than that of sentences like (6) to (12). But what I am arguing for is the thesis that insulationism is not appropriate everywhere, not the thesis that interactionism is appropriate everywhere. After all, we have in any case to allow some substantial degree of insulation between word-meanings in the composition of sentences. For example, if we want to make the sentence

The room is not square

able to function as the contradictory of

The room is square,

We cannot allow the subject of one such sentence to interact differently with its predicate than the subject of the other.

V

The conflict between insulational and interactional approaches to word-meaning has a profound impact on the analysis of metaphor. If the insulational approach is correct, and the meaning of a word or phrase in a particular sentence cannot be affected by the meanings of other words in the sentence, then *a fortiori* these other words cannot deprive a word or phrase of its literal meaning and force it into having a metaphorical meaning. According to the insulationist the phrase 'are made of pure gold' must have just the same meaning in

(18) The clouds are made of pure gold

as it has in

(19) The necklace is made of pure gold.

The insulationist must therefore give an account of metaphor that does not involve equating it with some special category of meaning, and if this cannot be done an interactionist account has to be preferred.

Davidson, for example, seeks[11] to treat metaphoricalness as what J. L. Austin would have called[12] a perlocutionary attribute of certain speech-acts rather than as a type of meaning that is sometimes exhibited by a sentence. On this view a metaphorical utterance has no content or sense other than its literal meaning. Instead it is regarded as a way of using a sentence with that meaning. I.e.. it is a species of language-use co-ordinate with, but different from, hinting, lying, promising, persuading, etc. So, according to the Davidsonian analysis, to paraphrase a metaphor is not to state its meaning but to evoke awareness of what the metaphor brings to our attention. But there are at least two good reasons why this analysis has to be rejected.

Consider, first, what happens in indirect discourse. Where an utterance is reported, the speech-act performed by the original speaker is not necessarily echoed by the speaker who reports him. Thus, if Tom utters the sentence

(20) I am sorry that I upset you,

he may well be apologising. But, if I then utter the sentence

(21) Tom said that he was sorry that he upset you,

I am not apologising myself: I am just reporting Tom's apology. The original speech-act is overridden by the passage from *oratio recta* to

oratio obliqua. Similarly, if I report a complaint, my report does not have to be a complaint. But metaphor behaves quite differently. When Tom describes his friend by saying

(22) The boy next door is a ball of fire,

Tom's intended description can be fully understood only by someone who understands the metaphor. But equally, if I myself report later

(23) Tom said that the boy next door is a ball of fire,

my report of Tom's utterance can be fully understood only by someone who understands the metaphor. The metaphor is not overridden by the passage from *oratio recta* to *oratio obliqua* and from one speaker to another. The oratio obliqua sentence contains the same potential for metaphorical meaning that the oratio recta contained. When my utterance of (23) reports Tom's utterance of (22), the phrase 'is a ball of fire' must be given the same metaphorical reading in my utterance of (23) as it would be given in Tom's utterance of (22). Arguably therefore metaphor inheres in sentence-readings, not in speech-acts.

Again, if metaphor is not a type of meaning that a linguistic expression may have but a way in which a linguistic expression may be used, it is impossible to explain how arguments can be constructed in which conditional statements have a metaphor occurring in their antecedents, as in

(24) If the boy next door were a ball of fire, he would not sleep every day till noon.

If the metaphor in (24) did not have the right kind of cognitive content, the clause including it could not function intelligibly as the antecedent of a conditional. For, whenever we want to investigate the truth-value of the conditional on a particular occasion, we need to paraphrase the meaning of the antecedent in terms of the vigour and alertness of the boy next door, not of his identity with a rare and highly dangerous atmospheric phenomenon. In other words the speech-act theory of metaphor falls victim to the same strategy of objection as that which has long since been shown to be fatal to the view that the function of 'This is good' is just to perform the speech-act of commendation or that the function of 'This is true' is just to perform the speech-act of endorsement.[13] The strategy is to make the sentence in question constitute the antecedent clause of a conditional, and then point out that the speech-act analysis is no longer applicable.

It may be tempting to argue that the grain of truth in the speech-act theory is that the possession of literal or metaphorical meaning is a

feature of an utterance, not of a sentence – or, if you prefer another way of putting it, a feature of a sentence-token rather than of a sentence-type. On this analysis, which Searle has proposed,[14] there is indeed a content that the message expresses and therefore a message that the metaphor can convey, though this content is not the outcome of compositional interaction between the words of a sentence-type. The content of the message is merely what the speaker means or intends by his utterance.

But here again difficulties arise. Consider a case in which the speaker has made a mistake in his choice of metaphor, analogously to the way in which people are sometimes guilty of malapropisms in relation to literal meaning. For example, despite her intention to speak about a nice arrangement of epithets the actual meaning of Mrs Malaprop's reference to a nice derangement of epitaphs was just what she said. To deal with such cases we distinguish the actual literal meaning of an utterance both from the literal meaning that the speaker intended it to have and also from the literal meaning that the hearer interpreted it as having. Ideally all three of these coincide, but in practice any of them can differ from the others. Nor is the range of possibility different for metaphorical meaning. Consider someone who believes – mistakenly – that Napoleon was finally defeated at the river Rubicon and therefore says, amid discourse about a friend's political career,

(25) Tom has met his Rubicon.

What is actually meant by (25) is that Tom has arrived at a limiting-point by passing which he becomes committed to a certain enterprise. But what the speaker intends to say is that Tom has met his Waterloo, and so what he intends to mean is that Tom has suffered a decisive defeat. So the intended meaning of the metaphor uttered is not the same as its actual meaning. Similarly, amid discourse about how the production of the new television series is proceeding, a speaker might say, by a slip of her metaphorical tongue,

(26) Bill is driving his own train.

when what she intends to say is that Bill is training his own driver. I.e. what the metaphor in her utterance actually means is that Bill is in executive command, when what she intends it to mean is that Bill is telling his deputy how to do the job. Of course, in normal communication it is generally more important for the hearer to grasp the intended meaning of a metaphorical utterance rather than the actual meaning (just as it is generally more important for the hearer to grasp the intended meaning of a literal utterance rather than its actual meaning). But that does not make the meaning of a metaphor depend necessarily

on the speaker's intentions. It is thus potentially misleading to describe metaphorical meaning as part of 'speaker's meaning', as Searle does,[15] since the speaker may not be correctly aware either of the literal or of the metaphorical meaning of what he or she says.

Nor would it help for 'speaker's meaning' to be understood here as what people normally intend to be understood by their utterance of a given expression, rather than what a particular speaker intends on a particular occasion. For then, because speakers of a natural language normally intend the sentences that they utter to have the literal meanings that those sentences actually have in their language, there would be no room for some intended meanings to be metaphorical.

We are left then with having to regard the metaphorical meaning of a sentence as determined by the literal meanings of the words that are used metaphorically and the interaction of these meanings with their linguistic setting and with relevant factual assumptions. Though there is a wide variety of ways in which this can occur, it is clear that the appropriateness of giving a metaphorical reading to a sentence – when it is appropriate to give one – must be marked by some intelligible sign. And that sign normally consists in a clash of some kind between the literal meaning of what is said and the presumptively known nature of that about which it is said. In other words, where the topic of a sentence is identifiable by conventional criteria, a metaphorical reading of the comment is prime facie appropriate if a literal reading of the comment would be grotesquely absurd. Moreover, the topic may be identified equally well by its being what the stressable grammatical subject denotes, like 'the clouds' in (18) and 'the boy next door' in (22), or by its being the underlying matter discussed, like Tom's political career in the discourse-sequence to which (25) may be supposed to belong. So in the interactions that promote a metaphorical reading topic dominates comment. Where comment dominates topic, as in (6)–(12), the interaction does not promote a metaphorical reading: otherwise it might be too difficult to identify the topic.

The overall picture of word-meanings that I am presupposing is one in which the fundamental, lexicographically recordable, default meanings of a word-type within a language are normally open-ended. The role of such a meaning in a particular sentence-reading – i.e. its contribution to that sentence-reading - may always be affected by the meanings of other words, phrases or clauses in it. In literal usage the dominating words, phrases or clauses help to generate relevant senses out of those allowable to the dominated ones in accordance with certain conventional principles, such as – at least in English – those concerning terms for representations, for users of skills, for types of things, and so on. In

metaphorical usage the general principle is that the dominating words or phrases (the ones that determine the topic) select relevant elements in what may be assumed to be sufficiently common knowledge about certain things, viz. those things that are denoted - in accordance with their conventional meanings - by the dominated words, phrases or clauses (which determine the comment). Thus in (18) what is selected is the known colour of the substance denoted by the phrase 'pure gold'; in (22) what is selected is the energy known to be present in fire-balls; and in (25) what is selected is the kind of part known to have been played by the river Rubicon in Julius Caesar's career.[16]

Note, however, that these items of knowledge must not be supposed to form part of the language-system. The analysis of metaphorical usage involves no such supposition, any more than the analysis of indexical usage involves supposing that what is referred to by particular utterances of 'now', 'this', 'he', etc. forms part of the language-system. What the language-system embraces, in each case, is a certain type of procedure for generating a sense or reference. The language-system does not embrace the actual sense or reference itself. Perhaps the closest comparison for the dependence of metaphor on presumed knowledge is with the way in which proper names function. The actual denotation of 'Tom' in a particular utterance of (23) or (25), for example, depends on the shared knowledge attributable to speaker and hearer of the utterance.[17]

Of course, the power of a metaphor is evident from the reception that is utterance evokes, and is normally the reason for using it. Some metaphors are apt, striking, or even inspiring, while others are obscure, feeble or pedestrian. But so too are different ways of saying the same thing without any use of metaphor at all. It is – *pace* Davidson – no less a mistake to treat the power of a metaphor as its primary linguistic function than to treat the perlocutionary force of an utterance as its meaning.

The meaning of a word or phrase, whether it be literal or metaphorical, may become clearer or more determinate as we embed the expression in longer and longer sentences. At a certain level of embedding the word, phrase or clause may be sufficiently disambiguated for normal purposes and the sentence can be uttered with the desired meaning in any, or almost any, situation. Alternatively some relevant features of the non-linguistic context, at the time of utterance, may be exploited for their ability to promote the required disambiguation. But this alternative is not an essentially pragmatic type of operation. If clauses or sentences describing the relevant features of the non-linguistic context are added to the linguistic context, the required meaning will be attributed by a hearer even if he or she has no other source of knowledge of the non-

linguistic context (such as perception of it). An example of this might be the way in which the meaning of 'lions' is determined by the addition of the relative clause in

(27) There are four lions here that are carved out of stone.

VI

What I have been arguing is that the interactionist conception of natural language definition is supported by a wide variety of linguistic intuitions, affecting both literal and metaphorical usage. In all the kinds of situation discussed, where sentences like (6)–(12), (19), or (22)–(24) are uttered, we seem to have sentences in which the meaning of one word or phrase is different from what it would otherwise have been, just because a certain other word or phrase is present. So on this view there may be interaction between the meanings of the sentence-parts that combine, in a conventional syntactic pattern, to make up the meaning of the sentence as a whole. And in order to elucidate how such interaction operates in a particular case we need to have not only definitions of the fundamental, lexicographically recordable, default meanings of the words concerned, but also a repertoire of semantic principles, like those concerning terms for representations, usable skills or types of things, or for dealing with metaphors *via* knowledge of the everyday facts that are presumptively familiar to most people within the speech-community.

There is a plausible objection that an insulationist account of natural-language definitions might be tempted to advance. If we suppose that the only meanings words can have in a sentence are what we would otherwise take as their fundamental, lexicographically recordable, default ones, then no interactions between word-meanings can take place at all, and the meaning of any sentence as a whole is generated directly by the fundamental meanings of its component words along with the syntactic and intonational pattern in which these are arranged. Instead of assuming the existence of interactions between fundamental word-meanings which would generate derivative word-meanings and thereby indirectly affect the meaning of the sentence as a whole, we have to suppose instead that the meaning of the sentence as a whole is affected directly by the co-presence of its words, phrases or clauses. Whatever semantic principles are needed for this have to be regarded as part of the compositional machinery that is in any case required in order to obtain the meaning of the sentence as a whole out of the fundamental meanings of its constituent words, morphemes, syntax and intonation. So, if (1), say, is transformed into (2), there is no change in the meaning of 'large' (which remains ambiguous), but there is a change in the meaning of the sentence as a whole. The issue here - the champion of insulationism might

insist – is not about choice of premisses for philosophical argument. Interactionists and insulationists can share premisses with one another by sharing the relevant linguistic intuitions about the fundamental meanings of the words in (1)–(3), (6)–(12), (18)–(19), or (22)–(24), and about the composed meaning of each sentence as a whole. All that they differ about is the structure of semantic explanation. The interactionist wants to posit derivative word-meanings, as well as fundamental ones, in order to explain how it is possible for natural-language sentences to have the meanings they do have, while the insulationist wants to avoid making this additional posit. So the insulationist's theory has the merit of being systematically more economical than the interactionist's. But the insulationist's theory does of course still differ here from the type of insulationism that fits a Tarski-type artificial language. Because it invokes, as part of the compositional mechanism, a variety of semantic principles that operate on fundamental definitions, it must admit, as against Montague,[18] that there are important theoretical differences between formal and natural language, even though both types of language favour insulationist definition. The part of ordinary English for which a Tarski-type language can offer a paradigm is substantially smaller than Davidson supposed.[19] It consists only of those literal sentences, if there are any, to which semantic principles about representational terms, etc. have no application.

VII

Despite this objection's prima facie plausibility some powerful considerations operate against it.

First, the interactionist account offers a general way of explaining why it is that certain sentences have the meanings they do. On the insulationist account each type of sentence like (1)–(3), (6)–(12), (18)–(19), or (22)–(24) has to compose its meaning through the operation of a separate semantic principle. On the interactionist account these semantic principles are all seen to share a common characteristic - the generation of derivative word-meanings - that explains the sentence-meanings in question. So interactionism offers a single, global, uniformity-based explanation, while insulationism offers a set of mutually independent types of explanation.

Secondly, the insulationist account leaves no room for words or phrases to be attributed metaphorical meanings. For on that account no derivative word-meanings at all enter into the explanation of how a sentence's compositional meaning arises. Instead, in appropriate cases, a sentence as a whole has to be attributed a metaphorical meaning. Yet in practice we may want to say that most of the sentence is not metaphorical at

all. And correspondingly the interactionist does allow us to distinguish within the meaning of a sentence as a whole those semantical components that are metaphorical and those that are not. So that in (24), for example, it is only the predicate of the antecedent that has a metaphorical meaning.

Thirdly, it needs to be remembered that there is another way of introducing non-standard meanings into a sentence. Instead of achieving this by exploiting the verbal make-up of the sentence, a speaker can preface or accompany the sentence by some appropriate stipulation. For example, he might stipulate that, whenever the name of an animal species occurs in what he follows, the name is to be understood derivatively as denoting a representation of a member of that species. But in such contexts the stipulation normally relates to a word or phrase, not to a sentence as a whole. Indeed, to stipulate how a sentence as a whole is to be understood seems to provide information that is more like a decode than a definition. So, if it is sentence-parts, not sentence-wholes, on which definitions are focussed, it should also be sentence-parts, not sentence-wholes, on which descriptive definitions of derivative meanings are focussed.

Admittedly an interactionist account of natural language does highlight a difference between Tarski-type formal languages and familiar natural languages. But perhaps it is right for this particular difference to be highlighted. Just as natural language exploits its indexical vocabulary in order to avoid having to assign unique names to each spatiotemporal position about which something may need to be said, so too it exploits its interactional principles in order to make it possible to say many more different kinds of thing with its descriptive vocabulary than a Tarski-type language with the same number of primitive predicates could make it possible to say. Each such principle that is accepted by a language enables it to dispense with an enormous number of definitions. That point lies at the heart of what is wrong with Davidson's semantic theory.

Notes

1. L. Jonathan Cohen, 'How is conceptual innovation possible?', *Erkenntnis* 25, 1986, pp. 223ff. The distinction has been fruitfully taken up by E. F. Kittay, *Metaphor: Its Cognitive Force and Linguistic Structure*, Oxford: Clarendon Press, 1987, esp. p. 110ff.

2. F. de Saussure, *Course in General Linguistics*, ed. C. Bally and A. Sechehaye, trans. W. Baskin, New York: Philosophical Library, 1959, 88-89. See L. J. Cohen, 'Chess as a Model for Language', *Philosophy* 11, 1982, 51-87.

3. See, for example, A. Garnham, *Psycholinguistics: Central Topics*, London: Methuen, 1985, 190-195.

4. J. D. Atlas, *Philosophy Without Ambiguity: A Logico-Linguistic Essay*, Oxford: Clarendon Press, 1989.

5. A. N. Whitehead and B. Russell, *Principia Mathematica*, Cambridge: Cambridge University Press, 1962.

6. See G. Frege, 'On Sense and Reference', in P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1952, 5678.

7. A. Church, 'A formulation of the logic of sense and denotation', in P. Henle, H. M. Kallen, and S. K. Langer (eds.), *Structure, Method and Meaning: Essays in Honour of Henry M. Sheffer*, New York: Liberal Arts Press, 1951, pp. 3-24.

8. Plato, *Republic* 597b, and Aristotle, *Categories ad init.*

9. *The Compact Edition of the Oxford English Dictionary*, Oxford: Oxford University Press, 1971, is somewhat inconsistent in its practice. For example: it lists representational senses for 'lion' and 'tiger' but not for 'giraffe' and 'zebra'.

10. C.L. Stevenson, *Ethics and Language*, New Haven: Yale University Press, 1944, and R.M. Hare, *The Language of Morals*, Oxford: Clarendon Press, 1952.

11. D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, 245-264.

12. J. L. Austin, *How to Do Thing with Words*, Oxford: Clarendon Press, 1962, 101ff.

13. See J.R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press, 1979, 92 123.

14. J. R. Searle, 'Metaphor', in A. Ortony (ed.), *Metaphor and Thought*, Cambridge: Cambridge University Press, 1979, 92-123.

15. Ibid., 100.

16. I no longer think it possible to give a distinctive-feature account of metaphorical meaning as proposed in L. J. Cohen, 'The Semantics of Metaphor', in A. Ortony (ed.), op. cit., 64-77. What inescapably block it are examples like (25), where indisputably factual, rather than linguistic, knowledge is invoked.

17. See L. J. Cohen, 'The Individuation of Proper Names' in Z. van Straaten (ed.), *Philosophical Subjects*, Oxford: Clarendon Press, 140-163.

18. R. Montague, *Formal Philosophy*, 1974, New Haven: Yale University Press, 188.

19. L. J. Cohen, 'A Problem about Ambiguity in Truth-functional Semantics', *Analysis* 45, 1985, 129-134.

9

A PROBLEM ABOUT AMBIGUITY IN TRUTH-THEORETICAL SEMANTICS

If one aims, as Davidson does, to build a semantics for a natural language on the foundation of Convention T, how far does one have to move away from the Tarskian paradigm in order to cope with the existence of ambiguity in many sentences of natural language? I shall argue that one needs to move a good deal further than Davidson supposes.

Obviously syntactic ambiguity, as in[1]

Someone heard the shooting of the hunters,

creates no special difficulties here. An adequate grammatical syntax will generate unambiguous structural descriptions, and under such a description a sentence like this can, as appropriate, either be said to be true if and only if someone heard the hunters shooting or be said to be true if and only if someone heard the hunters being shot. What precise form the postulates will need to have, in order that such T-sentences may be derivable, would obviously take quite a bit of working out, even where object- and meta-language are the same. But the extra dimension of complexity introduced into that problem by the ambiguity would call for a grammatical rather than a logical or philosophical resolution.

Verbal ambiguity, however, creates problems that are not so easy for philosophers to brush aside. Davidson's first attempt (in 'Truth and Meaning', *Synthese* 17, 1967, p. 316) to meet this problem was to argue that, if a verbally ambiguous sentence can preserve its ambiguity when translated into the metalanguage, 'a truth definition will not tell us any lies'. No doubt he had in mind such sentences as

(1) 'The box was in the pen' is true for an English-speaker at time t if and only if the box was in the pen before t.

And he claimed that in this respect ambiguity was just like vagueness or even like the property of being unsuitable for incorporation in a serious science. Where object- and metalanguage are the same, he said, 'all *these* problems will be translated without loss or gain into the metalanguage'.

Unfortunately vagueness and ambiguity cannot be treated alike in this regard. Consider for example the difference between

(2) George's time of arrival was fiveish in the afternoon,

and

(3) George arrived at 4.57 pm.

Obviously (2) is relatively vague, at least when compared with (3). And it would be quite appropriate to preserve this vagueness on the right-hand side of the relevant T-sentence. Sentences with relatively vague meanings are often needed for our self-expressive or communicative purposes, and statements of their truth-conditions must allow a correspondingly indeterminate spread of possibilities. So the sentence (2) is true if and only if George's time of arrival was fiveish in the afternoon, and to know that this is so might plausibly be said (at any rate from a realist point of view) to be part at least of what it is to know the meaning of (2). But the expression 'pen' is not a particularly vague one, at least in respect of the fact that it may signify either an instrument for writing with ink or an enclosure for ungulates or infants. Nor can knowledge of this ambiguity, which every competent English-speaker has, be usefully expressed in ambiguous terms, as in (1). An ambiguous description of some knowledge that a person has is not a scientifically satisfactory description of that knowledge, even if the knowledge in question is about an ambiguity. To assert such a description would be to make a pun, and play on words is scarcely a serious contribution to the investigation of linguistic competence.

It must have been for some such reason as this that three years later (in 'Semantics for Natural Language', in B. Visentini *et al.*, *Linguaggi nella Società e nella tecnica*, Milan, Edizioni di Communita, 1970, p. 182), without referring at all to his earlier account, Davidson proposed a quite different treatment of verbal ambiguity. Davidson argued that a sentence like 'The box was in the pen' created no insoluble problems for a theory of truth because it might be accommodated within a T-sentence like

(4) 'The box was in the pen' is true for an English-speaker x at time t if and only if either the box was in the playpen before t and the

> circumstances surrounding x at t meet condition c, or the box was in the writing-pen before t and the circumstances surrounding x at t meet condition c'.

Whether such a theory would then be acceptable, Davidson claimed, would depend on what we were prepared to allow in the description of the relevant conditions c and c'.

But some other issues also enter into the acceptability of the ensuing theory. Whatever form they take, conditions c and c' cannot jointly exhaust the possibilities. Some circumstances must be possible in which neither c nor c' prevails. Otherwise one would not be able to appreciate the inherent ambiguity of the sentence so well. And in such circumstances the predicate 'is a pen' must fail of satisfaction. Davidson's assignment of truth-conditions to verbally ambiguous sentences thus resembles, in one respect, Russell's assignment of truth-conditions to definite descriptions. Where none of the conditions disjunctively presupposed by the utterance of a certain verbally ambiguous predicate are actually present, the predication is false.

Perhaps it will be said that in constructing a semantics for natural language one need not be too worried about undisambiguated assertions of ambiguous sentences. Such assertions are deviant by the standards of ordinary conversational norms, so why not accept the convenience of being able to reckon them all false?

The answer to this objection is twofold. Where the ambiguity is a deliberate pun the point of the pun would be lost if the assertion had to be regarded as false. And where the ambiguity of the assertion is due to carelessness in matching sentences to circumstances appropriate for their utterance the author of the assertion is being convicted of a crime that he did not commit (viz. the crime of asserting what is not the case) and not convicted of the crime that he did commit (viz. speaking carelessly). Yet to prevent such anomalies it looks as if the semantics would have to admit the possibility of truth-value gaps.

To avoid having to make this admission it may be tempting to claim that verbal ambiguity is itself a rare, deviant and therefore quite unimportant occurrence in the sentences of a natural language. Homonyms like '(play) pen' and '(writing) pen' or '(legal) case' and '(suit) case' are not only very uncommon, it may be said, but also do nothing to improve a language's flexibility or intelligibility in everyday conversation. They are just accidents of etymological history. A semantics that was adequate for the rest of natural language would then be adequate enough.

But it is plainly wrong to suppose that the problem of verbal ambiguity in natural language can be satisfactorily resolved by tracing all

such ambiguity to the existence of admitted homonyms. The fact is that verbal polysemy, i.e. a word's having more than one sense, so far from being a rare and pathological phenomenon like homonymy, is the rule rather than the exception in natural language. Scan the pages of any English-English, English-French, French-English or French-French dictionary and the evidence lies in front of you. The words *not* assigned more than one standard sense are the odd ones: recent foreign imports, scientific coinings, thieves' argot, contemporary slang, and so on. The rest are polysemes. This is equally true of adjectives, as in 'healthy person' and 'healthy place', of nouns, as in 'table of mahogany' and 'table of figures', and of verbs, as in 'drops a bomb', 'drops a hint', 'drops a friend' or 'drops his voice'. Even prepositions have several familiar meanings: if George is 'in' the pool, the pool must be larger than George, but if George is 'in' trouble or 'in' the mood to go swimming, it does not follow that George's trouble or mood is comparable for size with George himself. Maybe there is a sense in which most of these phrases can be described as 'dead metaphor'. But if so most of the language we ordinarily use (including the phrase 'dead metaphor' itself) is dead metaphor. A semantics that ignores such polysemy thereby abandons any claim to be a semantics for natural language, because it ignores one of the main features – a feature comparable with indexicality – that differentiates a natural language from a Tarskian artificial one.

So the problem of verbal ambiguity cannot be swept under the carpet as the outcome of some pathological abnormality. Rather, it is an enormous convenience that the same word can often be uttered in one or other of several different though related senses. Instead of having to learn a very much larger number of words, each with fixed and context-independent meanings, we can learn a relatively small number of words with variable meanings and then exploit their verbal or situational contexts of utterance in order to disambiguate their actual occurrences. The economy of resources that is thus achieved by natural language is obviously comparable with the economies that are achieved by variability of reference in the case of proper nouns and other indexical expressions as an alternative to the fixed references available with the use of co-ordinate systems, individual constants, etc.

Moreover, if all ambiguous occurrences of such polysemes are to be dealt with as in (4), the question arises whether the disjunctions on the right-hand side of the ensuing T-sentences are always capable of being written down. Not only does the language need to be enormously rich in synonyms if this is to be possible, since so many potential disambiguations have to be expressible. But also the disjunctions have to be finite in length.

Can such finiteness be guaranteed? Consider the way in which a word that names a kind of object, situation or substance is also available to denote a two- or three-dimensional representation of an instance of this kind, as in

(5) Four lions surround the column.

Whether this usage exists in all natural languages, I do not know. But it is familiar in modern European languages, and its existence in ancient Greek was remarked by both Plato (*Republic* 597B) and Aristotle (*Categories*, ad init.). And the usage does not have to be learned separately for each noun denoting a species of animal, plant, artefact, event, etc. There may well be at least one such English noun that has never yet been uttered in a representational sense. Indeed that is presumably why the big Oxford dictionary (*The Compact Edition of the Oxford English Dictionary*, Oxford, Clarendon Press, 1984) does not list either of the words 'rabbit' or 'zebra' in this sense though it does list 'lion' and 'tiger'. But in relation to any such word we do not have to await an utterance of the relevant kind before being willing to accept that the representational sense is available in the language. In suitably disambiguating circumstances that sense is immediately understood. Nor is this limited to first-order representations. The word 'face', say, can not only come to mean – in appropriate circumstances – a representation of a face but also a representation of a representation of a face, and so on, as becomes apparent immediately one considers what one could say about a picture of a picture gallery in which one of the pictures is also a picture of a picture gallery. The semantic mechanism here is recursive, in that if a word can signify $\dot{x}$ it can also signify a representation of x. So it is arguable that (5) is, in principle, infinitely ambiguous. Or consider how the meaning of the word 'drop' changes from 'drop studying' in

(6) Most students here drop geography in their final year

to 'drop attending' in

(7) Most students here drop geography lectures in their final year,

and then to 'drop executing' in

(8) Most students here drop geography lectures reading assignments in their final year,

and then to 'drop paying' in

(9) Most students here drop geography lectures reading assignments library-fees in their final year.

If sentences can be of any finite length, there seems no limit to the variety of meanings that surrounding phrases can impose on words like 'drop', 'make', 'start', 'establish', etc. So

(10) Most students here drop them in their final year

looks infinitely ambiguous. Certainly the variations in the meaning of 'drop' in (6)-(9) cannot be put down to vagueness or generality rather than ambiguity. They cannot be regarded as meaning different species of a genus rather than as being different meanings of a word. This is evident from such sentences as

(11) John has dropped his geography lectures and Peter has done so also.

Here 'dropped' may imply 'dropped giving' in both cases, or it may imply 'dropped attending' in both cases. But it cannot, except as a rather poor joke, be used to refer to one kind of thing in John's case and to another in Peter's, though this would be quite acceptable in

(12) John has lost a pet and Peter has done so also,

where John may have lost a cat, say, and Peter a dog.

So it looks as though any T-sentence that stated the truth-conditions for (5) or (10) on the model of (4) would have to be infinitely long and the relevant satisfaction-conditions would neither be learnable nor statable in a finite period of time. Of course, it is easy enough to capture these infinities informally, by treating it as a semantic principle that, if a word can signify x it can also signify a representation of x, or that, if an agent does NP if and only if he or she VP's NP, then the agent drops NP if and only if he or she drops VP-ing NP. But, if a substantial number of such principles have to be grafted on to a truth-theoretical semantics in order to make sufficient provision for verbal ambiguity, the original felicity of the Tarskian paradigm is further compromised.[2]

Notes

1. For a defence of the existence of syntactic ambiguity here see A. M. Zwicky and J. M. Sadock, 'Ambiguity Tests and How to Fail Them', in J. P. Kimball (ed.), *Syntax and Semantics* Vol. 4, New York, Academic Press 1975, p. 13.

2. An earlier version of this article formed part of a paper presented at the May, 1984 conference on Davidson's work in New Brunswick, N.J. I am grateful for comments by several members of the audience on that occasion and for other comments by Jonathan Adler and Alan Weir.

10

THE INDIVIDUATION OF PROPER NAMES

Some of Sir Peter Strawson's most interesting work has been in connection with the ways in which certain features of natural language differ from the representations of them that are prevalent among formal logicians. The following is a latter-day effort that tends in a similar direction.

I

There are two different conceptions of a proper name. According to one, which I shall call the 'linguistic' conception, proper names are, and ought to be, individuated by their phonological form. Thus *Aristotle* is a different name from *Plato*, but the given name of the philosopher who tutored Alexander the Great is quite properly, and without any taint of equivocation, the same as that of the shipowner who married President Kennedy's widow. According to the other conception, which I shall call the 'idiosyncratic' one, proper names are, or would ideally be, individuated by their form-and-bearer. On this view it is either said that the philosopher's given name is different from that of the shipowner or, alternatively, that the form *Aristotle* is unfortunately at least two ways homonymous. The present paper is concerned to strengthen[1] the case for adopting the linguistic rather than the idiosyncratic conception, and to articulate some of the implications of so doing.

The idiosyncratic conception has been adopted by many modern logicians. Frege, like Russell, thought not only that in a logically perfect language every proper name would designate just one object, but also, presumably, that where a natural-language proper name designated now one, now another, object this was an imperfection and a kind of homonymy.[2] Quine wanted to replace *Socrates* by 'a general term that

is true of just one object'.[3] Though Dummett admits that the doctrine 'requires qualification in many respects before it becomes a realistic picture of our actual employment of names', he thinks that in principle 'a proper name possesses a unique and specific sense, common to all users of the name, which determines its reference'.[4] According to Kripke the main problem about proper names is whether such a name determines its characteristically unique reference by virtue of its sense, as Frege thought, or in some other way.[5] According to David Lewis 'we would distinguish the name of London (England) from the homonymous name of London (Ontario)'.[6] And within the framework of the idiosyncratic concept an immense amount of controversy has been generated by the question whether a natural-language proper name has a sense as well as a designation. On the one hand the assignment of a sense, or even of a disjunction of senses, seems to restrict the range of possibilities that can be considered. If *Aristotle* means the philosopher who tutored Alexander the Great, we cannot self-consistently consider what might have happened if Aristotle had not performed that task. On the other hand, if a proper name has no sense, it seems to have nothing with which to pick out a unique designation, in the way that the meaning of a common name like cat picks out the manifold extension of that name. So causal theories are constructed which tend, as soon as their details are made sufficiently determinate, to be too easily confounded by counter-examples.

The linguistic conception side-steps this tiresome controversy altogether, and guides us towards a less problematic semantical characterization. At least three considerations stand in its favour. The first two are easily stated: the third requires greater elaboration.

II

First, it is of fundamental importance to grasp that the properness of proper names is a feature–in Saussurean terms–of 'parole', not of 'langue'. In a particular utterance or sequence of utterances the name *Smith*, when functioning as a singular term, is typically peculiar to the person it names: it designates just one person. If more people of that name are at issue and the name is used without a distinguishing verbal or non-verbal context, then the use is ambiguous and deviant. But within any given natural language each proper name is normally available, on appropriately different occasions, for naming any number of entities of the appropriate category. During the course of one day a single speaker of English could unambiguously designate six different

people by the name *Smith* if he managed to have six conversations that differed appropriately in their circumstances.

Perhaps someone will object that we should think of each such conversation as taking place in an appropriate sublanguage of English, so that *Smith* can be conceived to have a unique designation in each sublanguage. But this objection cannot be sustained. No doubt it seems flattering: it makes brilliant linguists of us all. But the flattery is empty, because for many other theoretical purposes, whether synchronic or diachronic, we shall want a less extravagant method of individuating languages. A theory of language in such a Pickwickian sense of *language* would have rather poor explanatory value. It is methodologically unrewarding to have to sacrifice the unity of the larger entity (a language) in order to preserve that of the smaller one (a proper name's designation). Moreover, if you speak one language when you refer to your neighbour by *Smith* and a second when you refer to your colleague by that name, then what language are you speaking when they both come to your party and you proudly announce

> There are two Smiths here this evening,

or when, as does happen sometimes, you find occasion to refer separately to both of them in the same conversation? Someone may perhaps speculate that in the small groups in which he supposes language to have originated the philological ancestors of modern proper names had unique designations. But our task is to describe language as we know it is now, not as we speculate it to have been in the remote and unknowable past.

Convenience, not logical principle, is the main factor determining how many objects bear the same name. For example, the larger and better known the object, the more important that its name should not be widely shared. Only a few towns in the world are called London, but many English villages are called Coombe, and even more Englishmen are called John. Occasionally special scruples prevent reuse of a name, as in the case of Japanese emperors' names, but such scruples are religious or ethical rather than linguistic. It is misleading to describe this versatility of proper names as a form of homonymy, since all proper names are intrinsically versatile whereas homonymy, as in *bank* or *bat*, is an inconvenience from which almost all words of a particular grammatical category–almost all common-noun-forms, almost all verb-forms, almost all adjective-forms, etc.–are normally free. Admittedly languages sometimes extend the meanings of common nouns, verbs or adjectives, so that *blue* comes to signify a certain state of mind or a certain type of film, for example, as well as a certain colour. But this is an event in the history of the 'langue'–not an event in the history of 'parole',

like giving an old name to a new baby or a new ship: it is characteristically recorded by dictionaries, not by newspapers. The idiosyncratic conception of a proper name therefore carries with it the suggestion, as in Frege's writings, that natural-language proper names have drifted away from univocation, and deviated into homonymy, by some series of philological accidents or through the ignorance, laziness, carelessness, or improvidence of those who baptize children, name towns, etc. Yet in fact it is an immense convenience to have a reusable reservoir of phonologically acceptable speech-forms, divided perhaps into several categories, which parents of children, founders of cities or discoverers of mountains can draw on when they need.

Of course, we must distinguish between simple proper names and compound ones. The simple proper names of a language–sometimes called its proper nouns–are listed in dictionaries of male given names, female given names, surnames, field-names, place-names, etc. for that language. They are a species of word, co-ordinate with common nouns, verbs and the other traditional parts of speech, in any language in which these distinctions are syntactically important. On the other hand, the compound proper names of a language, which should perhaps be called its proper noun-phrases, are constructed out of its proper nouns or other words according to a variety of patterns. Thus we can have in English such compounds as *The Open University*, *Mount Everest*, *The New York Guaranty Trust Company* and *Charles Algernon Montgomery Smith.* But even compound proper names are reusable. Though it may be illegal for you to name your little school *The Open University*, you have not committed any linguistic solecism. Similarly, if Smith's parents failed to give him a sufficient number of personal names to distinguish him from all past, present, and future bearers of the same surname, they may have been deficient in parental care, but scarcely in linguistic propriety. And, after all, Smith's date and place of birth can always eke out identification when this does need to be unique, as on a passport.

Thus it is far from being the case, as the idiosyncratic conception implies, that a natural-language proper name normally has a unique designation, or that in an ideal natural language it would have one. Just the opposite. It would be a real imperfection in a natural language to prohibit the reuse of any proper name. Such a prohibition might be workable in a small, static community over a short period. But the phonological inventiveness it would require within any other kind of speech community would be quite beyond normal human powers. Just as language economizes a speaker's effort in saying different things by allowing an infinite variety of sentences to be composed out of a finite number of words, so too it economizes his effort in naming different objects by al-

lowing an infinite number of objects to be designatable (though not all at once) with a finite set of proper names. The idiosyncratic conception of proper names obstructs awareness of this immensely convenient feature of human language. If a logician's linguistic intuition generates that conception, rather than the linguistic one, he would do well to disregard his intuition.

III

The second consideration that stands in favour of the linguistic conception of proper names is the nature of human knowledge about them. That the philosopher who tutored Alexander was called Aristotle is normally accepted as being a matter of historical fact, not of language. People who do not know that philosopher's name, when asked, are taken to show historical ignorance, not linguistic imperfection, and the same is true if they do not know the name of a living person who is pointed out to them. Learning the given names of our new neighbours, or even their full names, does not necessarily add anything to our knowledge of English: we might well have known already that the names *John* and *Mary*–which our new neighbours happen to have–are English names. But when we learn our new neighbours' names we have at least then learned a fact that we need to know in order to avoid excessive formality in addressing them.

It is therefore important not to be misled by the partial parallelism between the function of a proper name in natural language and that of an individual constant in an interpreted formal system. If you don't know what the individual constant designates, you haven't got a complete knowledge of this artificial language. But you can have a complete knowledge of English, including even its whole present-day vocabulary of proper names, and still not know what given name or what surname–let alone what full name–any particular person has. Certainly there is a sense in which a particular individual constant in an interpreted formal system could have had a different designation. But to have given it that other designation would have constituted an alteration in the artificial language-system; so, relative to the original system, the constant could not have had a different designation from the one it in fact had. On the other hand our new neighbour's parents could perfectly well have decided not to call their son John, or could have changed the family surname, without changing any feature whatever of their natural language.[7] The name that he has is as much a contingent fact about him (changeable by deed poll) as the colour of his car (changeable by respray). Our language would be volatile indeed if every time anyone got married and

changed his or her name they thereby changed the language! Versatility, not volatility, is the hallmark of natural language.

It follows that where a mountain, a person or any other object has been given two different names, or where a person or corporation has adopted an alias, a statement asserting the identity of what the one name designates with what the other does is typically a contingent truth, not a necessary one. To describe it as a necessary, though a posteriori truth, would be to treat proper names in a natural language as if they were individual constants in an unalterable artificial language system. When we learn the two names we learn two facts about an object which need not stand in any form of necessary connection with one another. Of course, some cultures might have elaborate rules for naming children that generated a form of necessary truth for the utterance of an appropriate identity sentence in an appropriate context. Maybe in some remote Pacific island, peopled by English castaways, social conventions require that a man's name is John if and only if he is also called George. Once we know the rule we know, in a particular case, that John must be George. But the conceivability of such a bizarre culture serves merely to highlight the contingency of ordinary name-name identifications in natural language. If the ordinary ones were not contingent there would be no point in saying, on the Pacific island, that John must be George. More familiarly, perhaps, in the ancient Roman Republic anyone who bore the 'cognomen' *Cicero* was necessarily also named *Tullius*, since the Cicero family was part of the Tullian gens. But by these standards it is quite contingent for someone named *Nixon* to be also called *Richard*.

Philosophers who deny the contingency of any name-name identifications have tended to concentrate their attention on notoriously true identifications. Their thesis is intuitively less plausible when applied to false ones, like

(1) Bacon is identical with Shakespeare.

It scarcely seems impossible to imagine a course of events by which Nicholas Bacon's youngest son was born at Stratford and wrote the plays under a pseudonym as well as the *Novum Organum* under his own name. But these events never actually occurred, so (1) is contingently false. We therefore cannot say that if (1) were true it would be necessarily true: if (1) were true it would be contingently true. Equally, denials of true identity statements often seem eminently plausible candidates for contingent falsehood. If I say

(2) Hesperus is not identical with Phosphorus,

I may well be imagining a possible system of planetary revolutions that allows for this non-identity. Perhaps you wrote off your natural inclination to take

(3) Hesperus is identical with Phosphorus

as a contingent truth, on the ground that the apparent contingency of (3) is due to its announcement of an empirical discovery. But (2) announces no such discovery. Its assertion could well be a piece of unempirical dogmatism. So the natural inclination to take (2) as a contingent falsehood cannot so easily be written off; and if (2) is contingent so is (3).

It follows that there is bound to be a fallacy somewhere in any argument designed to show that all true name-name identifications are necessarily true. One such argument, for example,[8] proceeds from the premiss that everything has the property of being necessarily identical with itself. Hence George has the property of being necessarily identical with George. And if John is identical with George, John must have every property that George has. Therefore, it is claimed, John has the property of being necessarily identical with George. But the argument is essentially question-begging in that it assumes the sharing of properties, which goes along with identity, to include properties involving modalities. Or, alternatively, when the argument is set out formally, we can say that the fallacy consists in allowing modal operators to appear within substituends for predicate variables. Not that this procedure ought always to be formally impermissible. But it should not occur in a formal system that is designed to represent the logical structure of natural language.[9] Each syntactically differentiated kind of symbol-occurrence in such a formal system should be reserved, in the intended interpretation, for the performance of just one linguistic function, and just one kind of symbol-occurrence for each function: predicates for predicates, operators for operators, and so on. Go against this principle in your formalizations, if you please; but you are not then entitled to use your consequentially distorted representation of natural language as a basis for drawing related conclusions about what natural language is really like, as if you had not introduced any relevant distortions by your very system for representing its structure. After all, you would not judge your own appearance from its image in a distorting mirror.

Perhaps it will be objected that *necessarily* can occur within substituends for natural-language predicate variables if it occurs as a predicate modifier to express modality de re, and not as a sentential operator to express modality de dicto, and that its availability for use as such a predicate modifier is supported by the semantics for that use which has now been constructed.[10] But, in order to show that *necessarily* ac-

tually functions as a predicate modifier in relevant contexts in natural language, we need rather better reasons than the existence of this semantics. We need to be shown not merely what the semantics of such a modal predicate-modifier would be *if* the modifier were part of natural language. We need to be shown also *that* it is a relevant part of natural language. We need to be told of a statement like

Hesperus is necessarily identical with Phosphorus

that would have to be recognized both as true, and as having a thus modified predicate, even by those who adopt a semantics for proper names like the one detailed here: otherwise the objection would be question-begging. And the anti-essentialist has in effect always denied that any such statements exist.

IV

The third consideration in favour of the linguistic, rather than the idiosyncratic, concept of proper names is their actual mode of linguistic operation. In order to understand how proper names normally function in a natural language one has to grasp that this is a form of indexical, or token-reflexive usage. Logicians are ready enough to grant that demonstrative pronouns (like *he, outside, today, hereafter*), temporally tensed verbs (like *jumps* and *jumped*), or temporally oriented adjectives (like *present* and *future*) can have their designations determined, in actual speech, by their context of utterance. But they have generally been reluctant to recognize that this is how proper names normally function also.

Admittedly Tyler Burge[11] has recently argued that proper names function both as demonstratives (in *Alfred studies in Princeton*) and as predicates or general terms (in *There are relatively few Alfreds in Princeton*). But even this does not get it quite right. The concept of a demonstrative is a fairly clear-cut one, which loses a lot of its value if we blur it in order to include proper names within its extension. Demonstratives have their designations determined by their context of utterance without the benefit of any previous baptismal ceremonies, official registrations, etc. They are extremely adaptable tools of reference which a speaker can use without having any previous knowledge whatever about the entity designated (except perhaps knowledge of some highly general feature such as its sex). For example, the demonstrative *I* can be used in this way, without any impropriety or risk of misunderstanding, by a known amnesiac. But demonstratives pay the price for their extreme adaptability by being rather more difficult to use when the speaker is

at a distance–in time, space, or subject-matter–from what he wishes to designate. Demonstratives are typically helped to achieve their designation by their relative proximity, along one dimension or another, to what they designate: the easiest things to direct attention to, for purposes of mutual understanding, are what you and your addressee can both see or have just been discussing.[12] The use of a proper name, on the other hand, does not depend in any way on proximity to what it designates. No doubt proper names are substantially less flexible than demonstratives, and their optimally effective use requires both the speaker and his addressee to have stored in their memory certain appropriate items of non-linguistic knowledge. But proper names have the countervailing advantage that they can be used to designate entities which are far beyond the speaker's current range of pointing out. No useful purpose, therefore, can be served by confounding together two such valuably different categories of words. Even if we were to classify proper names as a species of 'demonstrative', in some extended sense of that word, we should still need another term to cover those 'demonstratives' that are not proper names.

Nor is it particularly significant, as Burge seems to imply, that proper names sometimes function as predicates or general terms, in sentences like

There are relatively few Alfreds in Princeton.

Burge objects to treating such uses as special ones, and wishes to regard modified and unmodified occurrences of proper names as co-ordinate types of usage. But Burge offers no evidence that the modified use of proper names occurs in most natural languages: all his examples are from English. And the fact is that all forms of indexical expression tend to lend themselves, at least in some languages, to de-indexicalization. This certainly happens with ordinary English demonstratives, though Burge seems unaware of it.[13] Compare, for example,

(4) That cat is a *he*[14]

or

(5) The *outside* world is unknown to us

or

(6) *Today*'s problem is always more exciting than *yesterday*'s

or

(7) Some people fear the *hereafter*.

And it also happens with temporally tensed verbs, as in

(8) Kittens *jump* when startled,

or with temporally oriented adjectives, as in

(9) *Present* love has *present* laughter.

In sum, ability to function as a descriptive term has not been established as a feature of proper names that distinguishes them from other indexically usable expressions;[15] and a technique for formalizing sentences that include proper names is obliged to represent a connection between their indexical and descriptive uses only so far as it also needs to represent a connection between sentences like

That cat is a he

and

He is purring.

Indeed, so far as it is the task of logical formalizations to represent such connections, one might think it more important to connect the designatory potential of proper names with their vocative use. Proper names are very commonly used to request or arouse the attention of the persons to whom they can, in those contexts, also be used to refer. Burge does not mention this use, though it is much more closely confined to proper names, among indexically usable expressions, than is deindexicalization: perhaps the second person pronoun is the only demonstrative that regularly functions thus. Yet any formalization of imperative sentences like

John, come here!

must be capable of representing vocatives.

V

How then does the context of a proper name's utterance determine its designation? We need to distinguish at the outset here between the object that a speaker intends to designate by his utterance of a proper name, the object that is actually designated, and the object that the addressee takes to be designated.

By a slip of the tongue, or as a result of originally mishearing, or later misremembering, the Peripatetic philosopher's name you might utter the sentence

(10) Aristophanes tutored a great conqueror.

You have it clearly in mind that Alexander was tutored by a philosopher, about whom you know a number of facts such as that he was born at Stagirus in 384 BC and walked about while lecturing; and you certainly do not think that Alexander was tutored by a very aged comic playwright. Whom does the name *Aristophanes* in your utterance designate? According to many logicians[16] it designates the person you had in mind or meant, viz. Aristotle. But 'designates' here is in any case a technical term and we have to adopt those criteria for its predication that are appropriate to the theoretical function of statements about designations. And these statements must at least give the facts that are needed for truth-assessments, if statements about truth-conditions form at least a part of any adequate semantics for a natural-language. Hence 'designates' must be so defined that what *Aristophanes* designates in your utterance of (10) must, if and only if your utterance is true, be an entity satisfying the predicate *tutored a great conqueror.* So we need to decide whether to call your utterance of (10) true, or false. Well, what did you say? You said that Aristophanes tutored a great conqueror. You may have intended to say that Aristotle did this, but what you actually said was that Aristophanes did it. No court of law, no jury of your colleagues, no verdict of public opinion, would describe what you said otherwise. Reformulation of your utterance in indirect speech supplies a decisive basis for truth assessment. After all this is the basis on which the biconditionals of truth-conditional semantics are constructed. The sentence 'Snow is white' is held to be true, for any speaker at any time, if and only if snow is white, just because, whatever such a speaker intends to say, what he actually says is that snow is white. It follows that, irrespective of your intentions or of your beliefs or of what you had in mind or of the causal origins of your mistake, your utterance of (10) was false. The name *Aristophanes* in your utterance did not designate the man whom you had in mind or meant. It is just the same with demonstrative pronouns. Suppose you say

(11) He is coming with me,

having John in mind and pointing over your shoulder, since you think that John is behind you; and suppose that Jim, not John, is the man behind you. Despite what you intended, what you have actually said is that Jim is coming with you and the demonstrative pronoun *he* in your utterance of (11) designated Jim, not John.[17] The objective designation of your pronoun depends not on what you yourself happened to believe, but on what an observer of your gesture might reasonably presume that you knew.

No doubt there is a sense in which you were referring to Aristotle, when you unfortunately used the wrong name for him, and to John, when you pointed in the wrong direction. But, even though it would be quite true, in one ordinary, everyday sense of *refer* to say this, that sense of *refer* is not the one needed to do the technical job for which I have been using the word *designate.* In the pragmatic, or speech-act, sense of *refer*, people and their utterances refer, and frequently what they refer to doesn't even exist, though they may think it does, like the book which Professor X referred to in his lectures and which turns out never to have been written. In the usual semantic, or logical, sense of *refer*, however–i.e. in the sense expressed by *designate* in the present paper–it is only words, phrases, individual constants, class-abstracts, etc. that refer, and existence is arguably a necessary condition for being referred to.[18] Again, in the everyday, pragmatic sense of *refer* you have not referred to John, but addressed him, when you say

John, come here!

Yet this need not prevent our saying that the first word of your utterance designates the subject of its imperative verb-form.

Very similar considerations refute other psychologizing accounts. For example, even though the person with whom you are talking is familiar with your tendency to say 'Aristophanes' when you mean 'Aristotle', it does not follow from that fact that the proper-name token in your utterance of (1) actually designates Aristotle. We have to distinguish here between the linguistic meaning of your utterance and the message that it successfully communicated.

How then is it determined what such or such an occurrence of a particular proper name actually designates? Four factors seem capable of playing a critical part, though they are not equally important in every utterance: the topic under discussion, the speaker's knowledge about what entities within that topic have the name in question, his addressee's knowledge about what entities within the topic have the name in question, and any relevant events, circumstances, etc. that are present at the time and place of the utterance and may be reasonably presumed to be known to the speaker and his addressee. In a book, for example, where the author of the text is rather remote from his readers and the readers may be rather heterogeneous in their education, the designation of a proper name will be primarily determined by the book's topic. In a history of logic *Aristotle* denotes the Stagirite: in memoirs of the twentieth century jet set it denotes the shipowner. In the course of a brief conversation, however, the topic of a remark may be insufficiently clear to fix the designation of a particular proper name, and the determin-

ing consideration may be the fact that only one person of that name is known to both the speaker and his addressee. Nor need they both have precisely the same knowledge. Maybe the speaker knows only that the founder of the Peripatetic school of philosophy had the name *Aristotle*, while his addressee knows only that the most famous person to be born in Stagirus had that name. But in virtue of knowing these two things they both know enough to establish the name's designation. Or perhaps the speaker and his addressee share knowledge of several people with the name in question, and the presumably observed arrival of one such person is the critical consideration. Of course, two or more of these factors may sometimes conflict with one another. But on those occasions the proper name's designation is ambiguous. And, if none of the four factors is appropriately operative, the designation is empty and the utterance correspondingly defective.

Note that it is the addressee who counts here, not the hearer. The background knowledge of casual eavesdroppers is irrelevant. But we can think of a normal bystander as adopting a neutral and independent position between the speaker and the addressee. He arbitrates on the actual, objective designation as distinct from the one subjective designation that is intended by the speaker or the other subjective designation that is imputed by the addressee. So any causal chains that have to connect the speaker's state of mind with the name's original assignment to its bearer must be matched by causal chains connecting the addressee's state of mind with this. Nor are these causal chains something special to proper-name designation. They are relevant here just because, first, all knowledge of fact entails some causal connection between the facts that are known and the belief implicit in the knowledge,[19] while, secondly, speaker's and addressee's knowledge about a name are two of the four factors that determine its designation.

Again, the uniquely identifying descriptions that some logicians have wished to associate with proper names do not belong to these names as such, any more than similar descriptions belong to demonstratives. But when we talk of the speaker's or hearer's knowledge that goes to determine a proper name's designation we must think of this as containing relevant descriptive items, such as that Aristotle was the founder of the Peripatetic school of philosophy or the most famous person born in Stagirus.

Thus it is excessively restrictive to suppose, as some logicians have done,[20] that a proper name achieves its designation of a particular object via the existence of a community in which it is common knowledge that members of the community have in their repertoire the procedure of using the name to refer to the object. There is no reason at all why

the use of a particular name to designate a particular object should not be private to a single speaker. For example, suppose a hermit calls his cat *Aristotle*. Then *Aristotle* is certainly the name of that cat, even if no one else ever knows this. The hermit has certainly succeeded in using a proper name to designate the cat without the existence of any community of the above-mentioned kind. To require the existence of such a community is to assimilate naming far too closely to meaning. It is the opposite error to that of those (perhaps legendary) philosophers who are supposed to have conceived all meaning on the model of naming–the notorious 'Fido'-Fido fallacy. Proper names are no more parasitic on the existence of such a community than are demonstratives. Some assignments of proper names are socially sanctioned and widely known, others are private to a few people, and yet others are quite individual and personal, like some love-names, nicknames, or pets' names. Similarly some occurrences of demonstratives achieve their designations via the use of gestures that enjoy a conventional significance within this or that sub-culture of a speech-community, while other occurrences of demonstratives, like most occurrences of the first person singular pronoun, need no such sub-cultural props. Certainly neither proper names nor demonstratives nor any other linguistic form could have designatory potential unless a community existed in which it was common knowledge that members of the community include the referential use of such sound-patterns within their communicational repertoire. But it is this general ability to use proper names referentially that needs to be included in a community's self-conscious repertoire, for just the same reasons as the other components of semantic competence need to be. Without such tacit conventions it is impossible to achieve communication by any system of meaningful signals, let alone by language.[21] Perhaps it will be objected that the hermit's name for his cat is not a real counter-example. 'His usage is parasitic', we may be told, 'upon acceptance of a standard procedure for naming in the community at large. The hermit just constitutes a one membered sub-community in which it is common knowledge that all members of the sub-community have in their repertoire the procedure of using the name to refer to the object.' Well, such an objection runs the risk of trivializing the thesis that it sets out to defend. We can agree that the hermit's usage is parasitic upon acceptance of a standard procedure for naming in the community at large. But it is not obviously parasitic upon acceptance that a particular name is not normally used in a sub-community unless it is common knowledge that all members of the sub-community have in their repertoire the procedure of using the name to refer to its bearer. And in any case consider a situation where more than one person is involved and

the speaker does not know what the addressee knows about the name in question. Suppose the topic of a guide's lecture is clearly ancient Greek culture, and the guide asserts (in any language) to his audience of exotic tourists

(12) Aristotle is a philosopher.

Such a speaker may well be the first person to inform that particular audience about the existence of a famous ancient Greek called *Aristotle.* But the name's designation in his remark remains unaffected thereby. Despite the historical ignorance of his audience, the guide must still be construed to have asserted that Aristotle (the ancient Greek) was a philosopher.

'That is beside the point', the objector may now claim. 'Nothing gets a name in the first place except by the existence of a community in which it is common knowledge that members of the community have in their repertoire the procedure of using the name to refer to its bearer.' But this is to look at the issue through the wrong end of the long telescope of time. What is at stake is not how an entity acquires its proper name, but how a particular occurrence of the name, perhaps centuries later, or continents away, obtains its designation. Some of the factors determining this designation do indeed relate to knowledge that a certain entity once acquired a certain name. But we must not confound the knowledge-conditions appropriate to name acquisition with the knowledge conditions appropriate to name-use. Ideally both speaker and addressee should know about the fact of name-acquisition, but neither need know that the other knows. A bystander could always determine the designation by looking only at the four factors already listed.

Finally we shall perhaps be told that, if the speaker wishes to use a proper name in a successful linguistic communication, he must know that the addressee knows the name's bearer, and that, if the addressee is to interpret the speaker's utterance correctly, he must know that the speaker knows the name's bearer: so in general there must be common knowledge that members of the community have in their repertoire the procedure of using the name to refer to the object. But one needs to distinguish here between the conditions requisite for a successful act of communication in which a proper name is employed, and the conditions determining what the proper name actually designates. It is just because the latter may include the first order knowledge-states of speaker and addressee that the former may have to include knowledge about these knowledge-states. To know what the name's occurrence designates, one must know what relevant conditions are present.

VI

The idiosyncratic conception of proper names, therefore, with its projected individuation of them by form-and-bearer, gains no support from considerations about the way in which natural language actually functions. Nevertheless the requirements of theoretical systematization often force us to oversimplify. Ideal gases, lines that have length without breadth, and perhaps also truth-functional connectives, are familiar examples of such well-motivated regimentation. They constitute misrepresentations only so far as we have some good reason to conceive our theories as descriptions, rather than idealizations, of their several domains. So logicians might be tempted to defend an idiosyncratic conception of proper names on the ground that it allows a very simple and straightforward procedure for the formalization of sentences containing proper names. Each proper name, when individuated by form-and-bearer, is assumed to correlate with a distinct individual constant in the formalism.

Now there is certainly no logical harm in this mode of formalization. It will not lead us from true premisses to false conclusions; and, given a sufficiently inexhaustible supply of individual constants, our formal language can represent thus (other things being equal) any argument that is constructable in natural language with the help of proper names or demonstratives. Indeed this mode of formalization is ideal for logical exercises and for the analysis of particular pieces of actual reasoning. But it will not serve as a satisfactory representation of how proper names function in natural language because, when operating as such a representation, it individuates proper names quite incorrectly and does not reveal the essentially indexical, or token-reflexive way in which their designations are determined. Nor, correspondingly, should this mode of formalization be used in the construction of a truth-conditional semantics for a natural language, if the theorems of that semantics are to state truth-conditions for sentences of the language that are identified and individuated by their appropriate structural description.

One method of dealing with proper names in a truth-conditional semantics, on a non-idiosyncratic basis, has been proposed by Burge.[22] It exploits the point that all English proper names have a descriptive or predicative use. His analysans for a sentence like (12) is a formal paraphrase of the sentence

> For all x and y, if x is an act of reference by a speaker s at time t to y with the first occurrence of *Aristotle* in *Aristotle is a philosopher*, then *Aristotle is a philosopher* is true with respect to s at t just in case the object which is y and is an Aristotle is a philosopher.

However, not only are Burge's proposals of dubious value for languages other than English. They also depend essentially on the mistaken assumption that what is relevant to the truth of an utterance is the speaker's subjective act of referring rather than the objective designation that a normal bystander would attribute to the utterance. Only on the former assumption would it be reasonable to restrict the relativization of a proper name's reference to relativization for a speaker-at-a-time. Nor, in any case, is there much illumination to be gained from an analysans which assumes that we already (prior to evaluating the analysis) know the conditions under which the utterance of a certain proper name by a speaker at time t is an act of reference to a certain entity.[23]

If instead we are right to suppose, for reasons already given, that we must represent the objective designation, we shall have to construct an analysans that increases the degree of relativization by acknowledging the four determining factors listed above. What a proper name designates in any given utterance will normally be relative to the subject-matter of the utterance, to the speaker's and addressee's knowledge, and to relevant features of the non-linguistic context. Thus the truth-conditions for a sentence like (12) will be given by a formal paraphrase of

(13) For all e, u, d, s, a and x, if e is the English sentence *Aristotle is a philosopher*, u is an utterance of e as an English sentence by speaker s to addressee a about domain of discourse d, and there is just one entity y such that y is identical with x and satisfies one or more of the following four conditions

- y has the name Aristotle and falls within d,
- s knows y to have the name *Aristotle* and to fall within d,
- a knows y to have the name Aristotle and to fall within d, or
- circumstances of u that are presumably known to s and a make it appropriate for u to include a mention of y,

then u is an utterance of a true saying if and only if x is a philosopher.

But the descriptive or predicative use of a proper name can be represented much more simply. For example, the truth condition for a sentence like

At least two Aristotles are philosophers

will be given by a formal paraphrase of

> For all x and y, if x and y both have the name *Aristotle*, then an utterance of the English sentence *At least two Aristotles are philosophers* is an utterance of a true saying if and only if x is not identical with y and both x and y are philosophers.

(I am assuming that two Aristotles here is synonymous with two people who each have the name 'Aristotle' and not with two members of the Aristotle family or with two people who resemble Aristotle.)

It is possible that I have underestimated the number and variety of factors that go to determine the objective designations of proper names. Perhaps the details of (13) need to be even more complicated. But the guiding principle on which (13) is constructed will still be unaffected. Proper names will just be even more token-reflexive than I have allowed for, and the idiosyncratic conception of them will be even more inappropriate than I have claimed in a semantics for natural language.

Perhaps it is also worth remarking that demonstratives and demonstrative phrases are not the only other types of expression for which, *mutatis mutandis*, an analysis like (13) is applicable. Titles (of books, plays, poems, stories, musical compositions, or works of visual art) may also be treated in this way. Titles may include proper names, as with *The Perils of Pauline*, but they need not, as with *Persuasion* and *Principia Mathematica*. They certainly differ from proper names in being substantially less adaptable than proper names as instruments of reference, just as proper names are themselves less adaptable than demonstratives. This is because titles, unlike proper names, create the presumption that they carry some information about what they entitle: *Hamlet*'s leading character is presumably a man of that name. But there is nothing in language to stop people from using the same title to designate different books, plays, paintings, etc. Title plagiarism may be a literary sin, just as company-name copying may be a legal wrong, but neither is a linguistic solecism. So a truth-conditional semantics for natural language has to relativize the designations of titles in much the same way as it relativizes those of proper names.

What should we say about Pegasus, Pickwick, 'et hoc genus omne'? Proper names may be given not only to actual persons, animals, places and events, but also to deities, angels and devils, to mythological heroes, to the persons and places of creative fiction, to military operations that may well be cancelled before they take place, or even to numbers that may one day be proved not to exist (like the lowest even number that is not the sum of two primes, which we might name *Eldorado*).

It is sometimes suggested that the problem of fictional reference may be solved by assuming that all references in works of fiction are pretended ones. What we should have to consider, on this view, are not

some special, fictional truth-conditions for the sentence *Spenlow came out of the house* that might be used to make the fictional reference to a Mr Spenlow, but just ordinary truth-conditions for a sentence asserting Charles Dicken's pretence that Spenlow came out of the house. But this suggestion will hardly work. Very few of the problematic references are pretended ones. Even in novels the author very rarely pretends to be referring to real people. To make such a pretence seriously, he needs to cite alleged sources or print photographs of mimeographed police dossiers, and so on. Most modern novels instead print an explicit disclaimer of any intention to refer to a real person, living or dead, so as to avoid the consequences of possible libel actions. One does not want to lose the distinction between fiction which pretends to be history and fiction which does not (or even denies that it is). In any case the generals who name an operation that they later have to cancel are not necessarily pretending anything when they refer to the operation by name in discussing its strategy and logistics: planning does not necessarily entail make-believe.

Indeed, one really should not be concerned, in an analysis of natural language, with the different domains of discourse to which speakers of the language devote their attention on different occasions. It is no business of linguistics to give some kind of priority to the physical over the spiritual, the historical over the fictitious, the achieved over the merely intended, or the confirmed over the merely hypothesized. Metaphysics should not be allowed to intrude into the semantics of natural language. Language is an all-purpose instrument and allows references to be made, in appropriate contexts, to the elements of any conceivable domain. So all that needs to be said about (13) in this connection is that the term 'entity' in it should be taken to permit a corresponding latitude. Since one of the four determining factors for a proper name's designation is the subject-matter of the utterance in which it occurs, we are entitled to suppose that the entity designated merely has to fall within the utterance's domain of discourse. Of course, occasions occur when even this requirement is not satisfied. Genuine mistakes are sometimes made as to whether a particular domain of discourse contains an entity bearing such-or-such a proper name. But it is of no relevance to the present enquiry whether one should adopt a Russellian or a quasi-Strawsonian method of dealing with those eventualities. In (13) truth conditions are given only for cases where the existential condition is satisfied; and acceptance of (13) leaves it an open question whether utterance of (12) should be regarded as an assertion of a false saying in the other cases, where the existential condition is not satisfied, or as an assertion of a saying with a neutral truth-value.

One final qualification needs to be added, however. Some languages divide their proper names into morphologically distinct categories, whereby people and places, for example, or men and women, can never bear the same names. Where this happens, the proper names have a good claim to be attributed corresponding meanings.[24] Thus in Latin the names *Julia* and *Julius* might be said to differ in meaning because their bearers must always differ in sex. But the point is easily represented within an analysans like (13). If English is such a language and *Aristotle* a male person's name within it, all that needs to be done to (13) is to replace 'entity' by 'male person'.

Notes

1. Some arguments in its favour were sketched in L. Jonathan Cohen, 'Searle's Theory of Speech Acts', *Philosophical Review*, lxxix, 1970, pp. 547-50. A further development of it was given by Tyler Burge, 'Reference and Proper Names', *Journal of Philosophy*, lxx, 1973, pp. 425-39 and by J. J. Katz, 'A Proper Theory of Names', *Philosophical Studies*, xxxi, 1977, pp. 1-80; but both Burge's and Katz's proposals are unsatisfactory in certain respects, which are discussed below.

2. Cf. 'On Sense and Reference', *passim*, in *Translations from the Writings of Gottlob Frege*, ed. P. Geach and M. Black, esp. p. 70.

3. W. V. O. Quine, *Word and Object*, 1960, p. 179.

4. M. Dummett, *Frege: Philosophy of Language*, 1973, p. 142. Cf. also 'What is a Theory of Meaning?' in *Mind and Language*, ed. S. Guttenplan, 1975, p. 132, where Dummett's argument against Davidson presupposes the validity of the idiosyncratic conception of proper names.

5. S. A. Kripke, 'Naming and Necessity', in *Semantics of Natural Language*, ed. D. Davidson and G. Harman, 1972, pp. 252 ff.

6. David Lewis, 'Truth in Fiction', *Amer. Philos. Quart.* V, 1978, p. 39.

7. This is implicitly denied by S. Kripke, 'Identity and Necessity', in M. K. Munitz, *Identity and Individuation*, p. 155, where he assumes that a renaming of planets entails a change of language.

8. Cf. S. Kripke, op. cit., pp. 135

9. The fallacy here is discussed at greater length, and in relation to other examples, in L. Jonathan Cohen, *The Implications of Induction*, 1970, pp. 211ff.

10. D. Wiggins, 'The *De Re* "Must": a Note on the Logical Form of Essentialist Claims' in *Truth and Meaning*, ed. G. Evans and J. McDowell, 1976, pp. 285 ff., and C. Peacocke, 'An Appendix to David Wiggins' "Note"', ibid., pp. 313 ff.

11. Op. cit. Burge's theory is endorsed by J. Hornsby, 'Singular Terms in Contexts of Propositional Attitude', *Mind*, lxxxvi, 1977, p. 42, and criticized (from another point of view than mine) by S. E. Boer, 'Proper Names as Predicates', *Philosophical Studies*, 27, 1975, pp. 389-400

12. In its original, now obsolete sense, the English verb to demonstrate meant to point out or indicate.

13. His treatment of ordinary demonstratives makes no mention of it: cf. T. Burge, 'Demonstrative Construction, Reference and Truth',*Journal of Philosophy*, lxxi, 1974, pp. 205 ff.

14. This usage has a paragraph to itself in the *Oxford English Dictionary*(Compact edition, 1971, vol. i, p. 1269).

15. Not that there are not other kinds of English word of which it seems more appropriate to say, as Burge says of proper names, that usage A is just an indexicalization of usage B, rather than that usage B is just a deindexicalization of usage A. The names for days of the week are a typical example, as in (usage A) *He will arrive on Sunday*, as distinct from (usage B) *Sundays are sacred.*

16. e.g. M. Devitt, 'Singular Terms',*Journal of philosophy*, lxxi, 1974, p. 188.

17. J. J. Katz, 'A Proper Theory of Names', *Philosophical Studies*, xxxi, 1977, pp. 1-80, has a linguistic token-reflexive theory of proper names, but spoils it by identifying the actual reference of a proper name's utterance with its intended reference.

18. On the difference between these two senses of refer, see further L. Jonathan Cohen, 'What is the ability to refer to things, as a constituent of a language-speaker's competence?', in R. Jakobson et al., *Linguaggi nella Società e nella Tecnica*, 1970, pp. 259 ff.

19. This is obviously too big a claim to be defended here. But so far as perceptual knowledge is concerned cf. L. Jonathan Cohen, 'The Causal Theory of Perception', *Proceedings of the Aristotelian Society*, Supp. Vol. li, 1977 pp. 127 ff.; and so far as non-perceptual knowledge is concerned this seems the only way to account for the knowledge of the examinee who mistakenly thinks that he is only guessing.

20. e.g. G. Evans, 'The Causal Theory of Names', *Proc. Arist. Soc.*, Supp. Vol. xlvii, 1973, p.202.

21. cf. L. Jonathan Cohen, *The Diversity of Meaning*, 1962, pp. 38 ff. The issue is investigated in detail in D. K. Lewis, *Convention: A Philosophical Study*, 1969.

22. 'Reference and Proper Names', *Journal of Philosophy*, lxx, 1973, p. 433. I am accepting for present purposes the general framework of truth-conditional semantics, as a way of talking about meanings in natural language, though in a more extended discussion this acceptance would have to be subject to some important qualifications.

23. A similar criticism may be made of the analysis of demonstratives suggested by D. Davidson, 'Truth and Meaning', *Synthese*, 17, 1967, pp.310-20, where Davidson claims that *That book was stolen* is true as (potentially) spoken by p at t if and only if the book demonstrated by p at t is stolen prior to t. A minor fault in Davidson's analysis is that according to it someone can know the semantics of English, in terms of the appropriate T-sentences, without knowing any relevant difference between *this* and *that.*

24. Exact translation of a proper name, from one language to another, may thus be rendered rather difficult. But in any case it is only by a rather crude convention that we assume the Stagirite to have been given the name *Aristotle* rather than its Greek original, which was a different speech-form and therefore a different name. The crudity of this convention generates serious problems for the analysis of indirect discourse, which are interestingly discussed in Saul Kripke, 'A Puzzle about Belief', in *Meaning and Use*, ed. A. Margalit (1979),pp. 239-83.

11

THIRD WORLD EPISTEMOLOGY

1.

Sir Karl Popper's striking hypothesis about a third world of objective knowledge deserves careful scrutiny because it is certainly his boldest conjecture. Moreover the hypothesis asserts an ontological thesis which can easily be adapted to fit certain non-Popperian methodologies, such as inductivism. So the hard core of the hypothesis is not hit by standard criticisms of Popperian methodology, and needs separate consideration. If it survives criticism, it will stand as a rather powerful theory about the nature of human culture: if it does not, then perhaps there is something to be learned from the reasons for its failure.

2.

In Popper's view the world of objective knowledge, as he calls it, or more generally the world of 'the objective spirit', is man-made[1]. But it exists to a large extent autonomously, because the moment our critical and creative thinking has produced a theory that theory may itself produce further problems. It follows that when any of these further problems is formulated by a human being, as in Goldbach's conjecture, the problem has been discovered, not made, by us[2]. Indeed a large part of the objective third world of actual and potential theories and books and arguments arises as an unintended by-product of the actually produced books and arguments and of human language[3]. There is thus a constant feed-back between ourselves and our work. It follows that the main aim of all historical understanding is the hypothetical reconstruction of a third-world problem situation. For example, how is Galileo's theory of

the tides to be understood? In order to achieve such under standing, says Popper, we have to investigate a fairly complex third world object[4]. The object consists not just of the standard problem about the cause of tidal motion, but of this together with its then current background of argument about the truth or falsity of Copernican theory, within a certain given theoretical framework of conservation laws.

It follows that the contents of the third world are rather diverse. Some, according to Popper, are just what Frege meant when he talked about the objective contents of human thoughts[5]. Indeed Popper tells us that 'theories, or propositions, or statements are the most important third-world linguistic entities'[6] But we are told also that the third world contains poetic thoughts as well as scientific ones[7]. Indeed it contains works of art, theoretical systems, problem situations, critical arguments, discussion-states, the contents of books and journals, and also scientific conjectures or hypotheses, experiments and their evaluations[8]. Moreover, just as the elements of the third world are thus richly diverse in character, so too are the processes in which they participate. There is not only a process of growth whereby a first problem gives way via tentative theorizing and error elimination to a second problem[9]. There is also competition and conflict between theories and problems, aspects of conjectures, interpretations and philosophical positions, and there are also comparisons, contrasts, analogies, problem-shifts, and so on[10]. Nor are all the objects in the third world of equal merit, since some objective knowledge is true, some false, some useful and some useless.[11] Indeed, while many theories in the third world are consistent, some – like Frege's *Grundgesetze* – are inconsistent, according to Popper[12]. As Popper puts it[13]:

> All work in science is work directed towards the growth of objective knowledge. We are workers who are adding to the growth of objective knowledge as masons work on a cathedral. Our work is fallible, like all human work. We constantly make mistakes, and there are objective standards of which we may fall short – standards of truth, content, validity, and others.

But the actual pattern of knowledge-growth is a Darwinian, evolutionary one. The fittest hypothesis survives[14].

Nor should the reader suppose that Popper is just indulging in an enormous metaphor or simile. He makes it quite clear by the comparison of his hypothesis with Plato's theory of forms, with Bolzano's theory of propositions in themselves, and with Frege's theory of objective thoughts, that he is ontologizing[15]. Indeed, he says in so many words that he is philosophically a 'realist' not only in regard to the physical world and the world of states of consciousness, but also in regard to the third world of objective knowledge[16]. His theory is meant to describe how knowledge really grows: it is not meant metaphorically.[17] Unfortunately, however, despite the length at which Popper has written about

his third-world hypothesis, a number of rather important issues remain unclarified.

3.

First, there is a problem about inter-communication. If intelligent creatures exist elsewhere in the universe with whom we terrestrial humans will never be able to communicate, the progress of objective knowledge may have gone much further than we shall ever know. If so, no human investigator will ever be able to test whether he has correctly conjectured the objective problem-situation in a particular area of enquiry at a particular time. Indeed, exactly the same problem arises in regard to any two cultures in the terrestrial past, if distance, xenophobia, geographical ignorance, or navigational incompetence prevented them from communicating information to one another. It follows that you cannot explain Galileo's theory of the tides, for example, by reference to the objective problem-situation in Galileo's time but only by reference to Galileo's subjective belief about that problem-situation. The explanatory goal of the third-world hypothesis seems unattainable.

One obvious solution of this difficulty would be for Popper to postulate a different world of objective knowledge for every closed group of intercommunicating intelligences. This would certainly deal with the problem across space, as it were. But, in view of the possibility that any such group of intercommunicating intelligences might be the outcome of past mergers or might give rise to future splits, Popper would also have to provide for corresponding processes of merging and splitting in his plurality of worlds of objective knowledge. And if the number of these worlds could thus fluctuate the conception of a third world of objective knowledge would hardly be on all fours with the conception of a physical world or of a world of states of consciousness. It is not now the conception of a totality of entities of a certain ontological category, but comparable rather to such less philosophical conceptions as that of 'the world of ancient China' or 'the world of the remoter galaxies'.

Accordingly, if the ontological unity of the third world is to be maintained, it would be necessary instead to suppose a potentially divided world, wherein perhaps many distinct cathedrals of objective knowledge were in process of mutually independent erection and were at different stages of growth. But the autonomy of the third world is now severely limited. Its ability to generate creative problems is now restricted by uncrossable frontiers between some of the objects that it contains. And these uncrossable frontiers are not the outcome or the explanation of any intrinsic epistemological factors. They are merely postulated ad

hoc as the counterparts of certain physical barriers between groups of intelligent creatures.

I do not know which of these two possible responses would seem least objectionable to Popper. But his third-world hypothesis can hardly come to grips with the problem of intercommunication without losing at least some of its initial simplicity and metaphysical glamour.

4.

There is a worse difficulty, however. 'Every scientific theory', says Popper, 'has, objectively, a huge set of important consequences, whether or not these have as yet been discovered'[18]. That is how the problem about Goldbach's conjecture came to exist in the third world long before it was actually discovered. And the third world also contains inconsistent theories, like Frege's *Grundgesetze*. But according to classical logic a contradiction implies anything whatever. So whatever proposition you care to mention is already derivable within some third-world theory. It looks as though the masons who are building the grand cathedral of objective knowledge have not only completed it but also rather disastrously over-fulfilled their norms.

How can this difficulty in Popper's hypothesis be circumvented? It is tempting to suppose that, in relation to every theory contained by it, the third world also contains the fact that the theory is consistent or in consistent, as the case may be. Indeed, Popper himself writes of 'the objective hitherto unnoticed fact, belonging to World 3, that there was an inconsistency in Frege's theory . . . '[19] After all, if the third world is attributed the capacity, as it were, to generate logical consequences, the capacity to detect inconsistencies might just as well be attributed to it. Then in accordance with such detections inconsistent theories can be consigned to some special category of demerit within the third world. But has the third world a solution for the decision-problem in regard to polyadic quantification-theory? No, for presumably Church's theorem is to be interpreted as a proof that the third world cannot contain a solution to this problem. It follows that the third world can hardly be supposed uniformly capable of including, whenever a new theory is fed into it, the fact that it is consistent or inconsistent, as the case may be; or we should be attributing to the third world a capacity that is demonstrably impossible.

It may seem preferable therefore to suppose that the third world operates with a relevance logic or with some other criterion of logical consequence that filters out any logical consequences that are due purely to the inconsistency of their premises. But any such supposition seems to

run counter to the basic conception of the third world as including *all* theories, since Gentzen's theory of natural deduction, for example, is a classical logic and its presence in the third world would seem to enable inconsistencies there to entail anything whatever. Nor, presumably, would Popper want to suppose that each theory in the third world is insulated somehow from every other, so that Gentzen's logic could never be used to deduce paradoxical consequences from Frege's *Grundgesetze.* Such a supposition would be tantamount to supposing that there are as many distinct third worlds as there are distinct theories. And it would also destroy the inner unity of the third world which is required to account for its autonomous creativity.

Nor would it help at all if Popper were to exclude internally inconsistent theories altogether from his third world. The trouble is that in order to remain sufficiently distinct from Popper's second world (the world of mind) the third world has to contain propositions and their interrelations, rather than assertions of these. Hence within the third world no distinction may be drawn between *my* assertion of a certain proposition or yours or Frege's. Though propositions or theories get there as a result of our intellectual activity in the first and second worlds, once they get into the third world they operate in their own right, as it were, and Popper insists that false propositions may be there as well as true ones. It follows that the third world may contain very many pairs of mutually inconsistent propositions supplied by different thinkers, and the co-presence of any such pair has presumably just as much right to be considered a conjunction (or a pair of premises for deduction) as the co-presence of any other propositions that strike sparks off one another within the autonomous creativity of the third world.

A plausible line of defence here might seem to be for Popper just to acknowledge that at any one time the third world might indeed contain a contradiction and consequently be over-full. He could then argue that from time to time the process of theory elimination would tend to cut it down to size. At any stage at which all inconsistencies had been eliminated the undesirable population explosion would be damped down. But this solution of the problem would leave us with the paradox that, each time an inconsistency returns to the third world as a result of human intellectual productivity, the third world reverts to an identical state of excess in regard to its population of propositions. Indeed, the third world can apparently be made to revert to that state even by sheer mischievousness. All you have to do is to contradict someone on some debatable issue, and thereby you sabotage the propositional selectivity of the third world until the debate is resolved.

In sum, when one presses the idea out into the requisite fine print, it is very difficult to see how Popper can maintain the autonomous creativity of the third world without getting involved in unacceptable paradoxes about the consequences of inconsistency. Yet without this autonomous creativity Popper's conception of a third world could make no claim to possess explanatory potential. It would merely gloss the actual record of second world developments.

5.

A further difficulty arises as a result of Popper's claim that the third world contains all the consequences of any scientific theory, whether or not these have as yet been discovered. For it seems that the third world may thus contain not only the problem for which Goldbach conjectured his solution but also the correct solution. If it is derivable from the postulates of number theory that every even number is the sum of two primes, or that some even number is not such a sum, then one or other of these two theorems ought already to be in the third world. It got there just as soon as the postulates did. Of course, the proof may not be there yet, and perhaps will never be. And perhaps, after all, Goldbach's theorem is neither provable nor disprovable. But clearly Popper's hypothesis entails that the third world contains an enormous amount of mathematical knowledge – in the form of provable but unproved theorems – which is as yet unknown to any human being. It follows inevitably that the growth of such knowledge cannot be a third-world process. In this respect at least the growth of human knowledge can only be represented by a growth of subjective aware ness, as human beings come gradually to discover provabilities that are already (according to Popper) part of objective knowledge. Correspondingly, the problem-situation at any particular time has to be viewed in subjective terms, not in objective ones, since it is only in human minds, not in the third world, that any of the relevant problems are as yet unsolved.

Nor is this just a difficulty for the Popperian epistemology of mathematics: it affects the history of natural science also. The apparent discrepancy between Newton's theory and Flamsteed's observations, in regard to the motion of the lunar apse, created an important problem-situation in early eighteenth-century mechanics[20]. Even Clairaut at first thought that the motion of the lunar apse was wholly inexplicable by Newton's laws, and started to formulate a new hypothesis regarding gravitation. But eventually he was able to carry out calculations showing that the observational results were indeed predictable from Newton's laws. Now according to Popper's hypothesis, if Newton's laws, plus the

relevant statements of initial conditions, were already in the third world when Clairaut began his investigations, then so too was the fact that the appropriate motion of the lunar apse is derivable from these premises. So, if the problem-situation that Clairaut encountered had really been determined by the third-world situation, he would never have wasted any time on trying to develop a rival hypothesis to Newton's and would instead have concentrated on his new calculations from the start. The only way for a historian to understand what actually happened at first is by reference to Clairaut's subjective but mistaken belief about the problem-situation.

It would be open to Popper to exclude this kind of difficulty in this third-world hypothesis by revising his view so as to deny that the logical consequences of a theory enter the third world before some human scientist discerns them. But then the autonomy of the third world would be gravely diminished and thereby, as remarked earlier, its claim to possess explanatory potential would be undermined. The inconsistency in Frege's *Grundgesetze*, for example, would not have entered the third world until Russell discovered it. So it wouldn't have been the objective autonomy of the third world that threw up the inconsistency, as it were, but the subjective creativeness of Frege's mind. Equally, Goldbach would have to be given the credit (which he surely deserves) for having posed the question that his conjecture was designed to answer. All that would have pre-existed Goldbach's question was the terminology in which it was formulated, along with some other (presumably suggestive) questions about similar topics.

6.

It might be urged, on behalf of some supposed nugget of truth in Popper's hypothesis, that even when one speaks of, say, Clairaut's mistaken belief about the problem-situation, one assumes the existence of an objective problem-situation for Clairaut to misconceive. But if the third world is to contain objective problem-situations, as distinct from misconceptions of them, it must presumably contain at any one time the currently best possible evaluations of whatever theories are under consideration at the time. The comparisons and appraisals that enter into a particular problem-situation must be those that ideally rational scientists would make, on the basis of available evidence, rather than those that fallible humans like Clairaut would make. Presumably no detectable errors would enter the third world. But then Frege's *Grundgesetze*, for example, would have been barred at the outset because of its inconsistency, contrary to Popper's express stipulation, and there would

be no explanation, within the third world, for Russell's discovery. A history of the third world would be like that kind of misleading history of science which skips blithely from one peak of success to another and overlooks most of the errors and half-truths in the valleys below. (Some related difficulties about the connections between Popper's three worlds are discussed by Currie [1978].)

7.

Finally there is serious difficulty about Popper's claim that his epistemology is 'a largely Darwinian theory of the growth of knowledge'[21]. On this view theories compete for survival and the fittest survives.

But which are the species and which are the species-members in this epistemological Darwinianism? If theories are cast in the role of species, then presumably the species-members are the representations of these theories in the minds and writings of individual scientists. But if that is so then the real struggle for survival takes place in the first and second worlds, not in the third. And yet a refuted theory is not automatically deleted from all the books in which it was printed. Many tokens of it remain, even when its more successful rivals enjoy much greater popularity. It follows that competition between theories cannot be envisaged as a Darwinian struggle in which theories operate as species and theory-representations as species-members. Are, then, the theories themselves to be conceived of as species-members, competing for individual survival in the third world? But if so, they seem for the most part rather ill-suited to constitute species. They serve their intended purposes only in virtue of being distinctively tailor-made for these and therefore different from one another. In sum, one is inevitably driven to the conclusion that any talk about Darwinian evolution in Popper's third world is either false or just metaphorical, contrary to Popper's own claim[22]. The complaint here is not that developments in Popper's third world are not the result of genetic mutations or physiological adaptations. How could they be? The trouble is rather that the central feature of the evolutionary process has no counterpart in the third world. Speciation does not play any important role in third-world development. Yet it was what Darwin wrote his great book about.

8.

Actually, the situation is even worse than this. The proposed analogy between Darwinian evolution and the progress of knowledge does not break down merely over the question of speciation. It suffers from several other flaws. Nor should this be at all surprising. Darwin's theory was

originally put forward as an alternative to those theories which sought to explain adaptedness by reference to some kind of rational design. Instead of appealing, like Christian theology, to God's rational plan, it contrived a pattern of explanation, via natural selection among random mutations, which had no need to assume the occurrence of any rational act. And not only was Divine rationality – i.e. the application of Divine knowledge for the fulfilment of Divine purposes – irrelevant to Darwinian explanation: no other kind of rationality was allowed to have a part in it either. In particular, no heritable innovations were to be traced back in Lamarckian style to learned adaptations. So no evolutionary change of any kind came about through the application of intelligence and knowledge to the solution of a problem. But by these standards science, in striking contrast, is a veritable paradigm of rationality. Scientists use their intelligence and previously acquired knowledge to discover more about how things happen, and they often learn thereby also how to prevent things from happening. Every stage in the progress of science is the product of intellectual effort and builds consciously on previous intellectual achievement. It seems quite inappropriate therefore, if not deliberately perverse, to think of this progress as a species of Darwinian evolution – as if something that is essentially rational could be a species of some thing that is essentially non-rational.

Of course, if you ignore the conception of rationality that contrasts with natural selection and suppose that rationality operates only via mechanical heuristics, then even science, or most of it, will seem non-rational. There is certainly no mechanical procedure for discovering explanatory hypotheses in unfamiliar fields of investigation. The main rationality of science consists, however, not in the partly unconscious processes that result in the production of a hypothesis or family of hypotheses but rather in the conscious, intelligent and well-informed application of appropriate criteria for selecting problems, appraising experimental designs, and evaluating hypotheses. Correspondingly, natural selection is a non-rational process because it involves neither consciousness nor intelligence nor the application of knowledge. That was at the heart of Darwin's idea. That was what shook the nineteenth century theologians. That is what sets Darwinian evolution utterly and irremediably apart from the rational processes involved in scientific enquiry. So any attempt to construe science as an evolutionary process can be expected to founder in a sea of incongruities.

We have already seen that one of these incongruities is the fact that science contains no obvious analogue for evolutionary speciation in nature. In terms of Peirce's type-token distinction, it is theory-types that get attacked by counter-evidence, not theory-tokens. Yet in nature it is

the individual members of a species that get wiped out by ecological hazards. So, in principle, one refutation of a hypothesis may be sufficient, so long as it is replicable, to demote the hypothesis within Popper's third world. But many individual members of a species need to be prevented from reproducing themselves if the species is to be wiped out. Again, it is perhaps the ideas and principles constituting a theory that are supposed to correspond to the phenotypes instantiated by a biological population. But what corresponds to the gene-pool then? Or if the ideas and principles are supposed to correspond to the gene-pool, and some general characteristics of theories are supposed to correspond to biological phenotypes, then why is it that the marriage of two mutually consistent theories produces a quite determinate offspring? That is, exactly the opposite happens to what the comparison with genes suggests: the set-up is deterministic, not probabilistic.

A further pair of incongruities arise in relation to the proposal of hypotheses.

First, the nature of the hypotheses proposed is strongly affected by factors of the kind that also control preferences between hypotheses. A scientist who has read up the literature in his chosen field of research has a mind that is well-stocked with factors that will affect the content of the hypotheses that occur to him, and at least some of these factors - in particular, beliefs about previous experimental results - will also be relevant to his evaluation of those hypotheses. The evolutionary analogue would be for the nature of genetic mutations to be causally affected by the kind of environmental pressures that are operating. But any such coupling is excluded by evolutionary theory. It may have a place in Lamarckian theory, but not in Darwinian. Similarly, cross-lineage borrowing is not an evolutionary mechanism, but scientists in one field often exploit experimental techniques or mathematical procedures that have been developed in another field.

Secondly, if scientists actually tried to take survival value into account when they put forward their hypotheses, they would aim at hypotheses which were very difficult to falsify. Cowardice, not boldness, would bring rewards. If instead they followed Popper's advice and aimed at hypotheses which were in principle highly falsifiable, they would be doing exactly the opposite of what an evolutionary interpretation of science requires. In particular, if their hypotheses should normally generate novel kinds of testable prediction, as Popper – like Bacon – recommends, this would have to be interpreted in terms of the evolutionary analogy as an unnecessary courting of danger from environmental hazards. Popper's falsificationism is deeply at variance with his evolutionism. Perhaps it will be said that evolutionism here is to be construed simply as an account

of what does happen in the progress of science, not of what ought to happen: it does not make normative recommendations to scientists and therefore what it says cannot be in conflict with any recommendations. But even if we admit this it will not eliminate the incoherence from Popper's philosophy, since his falsificationism is not just a normative recommendation. Falsificationism is supposed also to represent what the mainstream of science has always been like.

Finally, one might well ask, in regard to Popper's evolutionary interpretation: what new research into the history or sociology of science has it stimulated or could it stimulate? The answer seems to be: none. The actual similarities of structure between scientific progress and biological evolution are so few and so slight, and are counterbalanced by so many important dissimilarities, that the evolutionary interpretation is obviously much more likely to befog the historian or sociologist than to shine a search-light ahead of him.

Of course, none of this in any way counts against the conception of human science as a factor in the biological struggle for survival. It is highly plausible that some kinds of deductive and inductive mechanisms, and some innate quality-space, are a part of humanity's genetic heritage. It is also at least plausible to hold that modern science constitutes the most effective method of putting that innate endowment to work in the pursuit of knowledge. But nothing follows from this that would allow us to treat the actual progress of science as an evolutionary process. No *natural* selection operates, so far as we can tell, to bring about changes in scientific procedure. The time-scale is grossly inappropriate. Science can make vast progress within a single year of a man's life, whereas any evolutionary changes in human biology would require several generations to become manifest. Nor is it even clear yet that, in the long run, science actually has got survival value. Perhaps a species' possession of a certain kind of innate cognitive endowment in evitably leads eventually to the development of nuclear physics, and this – given normal animal aggressive instincts – eventually leads a species to self-destruction. Ipso facto, however, the human species may well be unable to survive long enough for any of its members to know for certain either that this is the case, or that it is not the case. So even at this quite general level nothing definite seems to be gained by

thinking of science within the conceptual framework of Darwinian theory.

9.

We are therefore driven to reject Popper's hypothesis about a third world of objective mind in Darwinian evolution. Critical analysis reveals far too many paradoxes and incongruities for it to afford an acceptable theory of cultural development. Somehow this development has to be accommodated within a more restricted (monist or dualist) ontology[23].

Notes

1. Popper, 1972, p. 147.
2. *Ibid.*,pp.l60-161.
3. *Ibid.*, p. 117
4. *Ibid.*, p. 170 175
5. *Ibid.*, p. 109.
6. *Ibid.*, p. 157
7. *Ibid.*, p. 106
8. *Ibid.*, p. 106 111
9. *Ibid.*, p. 121
10. *Ibid.*, p. 165
11. *Ibid.*, p. 115
12. *Ibid.*, p. 57
13. *Ibid.*, p. 121
14. *Ibid.*, p. 261 264
15. *Ibid.*, p. 106
16. *Ibid.*, p. 107
17. *Ibid.*, p. 261
18. Popper and Eccles, 1977, p. 140.
19. *Ibid.*, p. 57
20. Cf. Cajori, 1962, pp. 648 650.
21. Popper, 1972, p. 261
22. *Ibid.*, p. 261
23. This is a revised and expanded version of a paper which originally appeared in *The British Journal for the Philosophy of Science*, 31 [1980], pp. 175–180.

References

Cajori, F. (1962). Appendix. In Cajori, F., editor, *Sir ISaac Newton's Mathematical Principles of Natural Philosophy and his System of the World.* University of California Press, Berkeley.

Popper, K. (1972). *Objective Knowledge: an Evolutionary Approach.* Clarendon Press, Oxford.

Popper, K. and Eccles, J. (1977). *The Self and Its Brain.* Springer Verlag, Berlin.

12

GUESSING

1. The epistemological relevance of the problem

In recent epistemology the concept of guesswork or conjecture has been much used, though seldom, if ever, analysed. Its use has taken two main forms.

Sometimes the concept is used to formulate some quasi-paradoxical thesis in the philosophy of science. "The actual procedure of science", writes Popper (*Conjectures and Refutations*, 1963, p.53) "is to operate with conjectures: to jump to conclusions often after one single observation." Popper thinks of this view as being fundamentally opposed to the'inductivist','justificationist' philosophy which he attributes to Carnap. But Carnap too once wrote: "All knowledge is basically guessing" (*Mind, Matter and Method: Essays for Feigl*, ed. P. K. Feyerabend, 1966, p.249). In this way much of modern scepticism in the philosophy of science takes a less passive point of view than did the earlier scepticism of Montaigne and Hume. It holds that rationally sophisticated enquirers ought to engage in the activity of guesswork, rather than that they ought to relapse into the passivity of doubt.

The other way in which the concept of guessing has come to the fore is through the formulation of arguments for formal or informal theses in epistemic logic. It is disputed, for example, whether or not 'x knows that p' implies 'x believes that p'. Those who wish to reject this thesis sometimes seek to suggest examples of situations in which we should describe a person as both knowing that p and also thinking that he's merely guessing that p, since it is assumed impossible for the same man at the same time both to think that he is merely guessing that p and

also to believe that p. A typical suggestion of this kind is Radford's example of the French Canadian being quizzed about the dates of Tudor and Stuart monarchs (in "Knowledge by Examples", *Analysis* xxvii, 1966, p.4). The examinee answers a series of inter-connected questions correctly, but with obvious signs of doubt and hesitation. So he is said to know that p, q, r, etc. without believing these things, on the ground that he obviously thinks he's guessing.

Now what is a guess? How does guessing differ from, or resemble, conjecturing, speculating, suspecting, assuming, supposing, hypothesising and other kindred forms of conduct or states of mind? Within what limits, if any, can we pursue these various activities, and what does their pursuit exclude, or imply, about other matters? Curiously, there are hardly any discussions of these problems in the literature of modern epistemology. The concept of guessing has been often used but never analysed with any degree of thoroughness. But that really will not do, if the role of guesswork in science and its relationship to knowing and believing has to be determined with any degree of accuracy and rigour. The concept-cluster of guessing, conjecturing, speculating, suspecting, assuming, supposing, hypothesising, *etc.* demands elucidation as an essential preliminary to the construction of any theory about the more highly valued cognitive states or activities, like knowing, recognising, discovering, explaining, understanding, etc. So the task on which I am engaged is a rather modest and pedestrian one. There will be no opportunity for deep insights or grand theory-construction. But the preliminary cleansing of one's apparatus is just as important, *mutatis mutandis*, in the philosophy of science as it is in experimental science itself, and unless we have a reflective, self-conscious awareness of what guessing is we shall be unable to appraise accurately its relation to other, and perhaps more important, intellectual activities.

I should make one thing clear at the outset. I am not concerned at all with the mock-modest use of 'I guess'–a use commoner now in North America than in England, where it originated. In that use the sentence

It's safe here, I guess

expresses not a mere conjecture but a well-grounded belief and its utterance can give considerable reassurance. Guessing, in that use of the word, is not guesswork. But I shall say no more about that use since my concern is wholly with the sense in which it is tautologously true to say that to guess is to proceed by guesswork.

2. Guessing, conjecturing and speculating

Consider first how guessing compares with conjecturing.

On a superficial glance it may be tempting to suppose that the English verb 'to conjecture' is just a polysyllabic Latin-born word for precisely the same type of act as is denoted by the monosyllabic Anglo-Saxon-born verb 'to guess'. It would follow, if this were true, that the only semantic difference between the two words would be the usual connotation of snob-value, or social respectability, that goes with a Latin as distinct from an Anglo-Saxon etymology. Synonym-pairs of this kind are particularly familiar in the vocabulary of human action. The genteel physician speaks of 'intercourse' and 'defecation', not of their Anglo-Saxon equivalents, and in military jargon an army unit is ordered to 'proceed' to a destination, not to 'go' there. So to describe something as a conjecture, it might be thought, is just to impute rather more respectability to it than to describe it as a guess. Notoriously, mathematicians sometimes record what they call 'conjectures' in professional journals, but guesses, it seems, are not so honoured. Popper called his book *Conjectures and Refutations*, not *Guesses and Refutations*. On the other hand, if someone comes up behind your back and puts his hands avert your eyes, he doesn't say

> Conjecture who!

he says

> Guess who!

There are guessing games, but not conjecturing games. In short, it seems, children make guesses and so do many gamblers, professors and research-workers make conjectures.

However, there is considerably more to it than this. A distinction like that between guessing and conjecturing is found also in languages without a pattern of etymological differences corresponding to the Latin-Anglo-Saxon difference in English *e.g. raten* or *erraten* in German, as against *vermuten* or *mutmassen*, or *deviner* and *conjecturer* in French. Moreover, not all the differences between guessing and conjecturing are differences of intellectual repectability. Both guesses and conjectures may turn out to have been good or bad, correct or incorrect. But there is no sense of the verb 'conjecture' in which

> John conjectures that p

implies that p. 'Conjecture' never seems to behave as a success-verb, like 'discover', 'know', or 'remember'. On the other hand 'guess' has both a neutral and a success sense. We can say not only

> She guessed that it was *Tom* at the door, though in fact it was *Bill*

but also, with a different stress-pattern,

> She *guessed* that it was Bill at the door, though it might well have been Tom.

Correspondingly a man may guess the riddles and puzzles that are posed to him, whereas to conjecture a puzzle is to put a problem forward rather than to succeed in solving it.

Again, if a man conjectures that p, he may equally well be said to have conjectured the hypothesis that p. But the phrase 'the hypothesis' does not afford an internal accusative to 'guess' as it does to 'conjecture'. The sentence

> He guessed the hypothesis that p,

if it means anything at all, means more or less the same as

> He guessed that the hypothesis that p was the hypothesis wanted

rather than the same as

> He guessed that p.

Correspondingly, to report someone as having conjectured that p is normally to raise a rather more restricted range of questions than to report him as having guessed that p. Certainly the question will normally arise as to whether the hypothesis conjectured has been verified or falsified, proved or disproved, confirmed or undermined, etc. But if a man is reported as having guessed that p, the question raised is just as much about the causes, effects, manifestations, etc. of the guess as about its correctness or incorrectness. Hence we say

> His guess is that the President will sign a peace-treaty with Hanoi to-night, so at present he's selling his shares in Dow Chemicals

rather than

> His conjecture is that the President will sign a peace-treaty with Hanoi to-night, so at present he's selling his shares in Dow Chemicals.

But, though the latter sentence does not sound quite as unproblematic as the former, there is nothing at all odd in saying

> His conjecture is that the President will sign a peace-treaty with Hanoi to-night, so at present he's checking the whereabouts of the Hanoi delegates.

Talk about conjectures thus concentrates attention on to problems about correctness in a way that talk about guesses does not. In reporting an individual's conjecture that p a speaker suggests some fairly general context of ignorance whether p. But in reporting someone's guess that p he makes no such suggestion. And this is all quite in keeping with the two differences already noted between the verb 'to conjecture' and the verb 'to guess', *viz.*, that the former but not the latter takes the phrase 'the hypothesis' as an internal accusative and that the latter but not the former sometimes itself implies the correctness of what is guessed. The word 'conjecture' is therefore peculiarly well suited to the task of describing the serious, high-level activity of mathematicians, scientists and other types of investigator, while the word 'guess' is better suited to describe what happens in certain kinds of game, where just one person has to guess and the others already know, or could know, the answer.

Finally, these differences seem to be connected with an important underlying fact. Admittedly the two English verbs 'conjecture' and 'guess', like the more or less corresponding pairs of words in French, German, *etc.*, are often used indistinguishably. But there is also an important difference that they can be used to mark, *viz.*, the difference between forming an opinion on admittedly inadequate though not wholly nonexistent grounds and forming it on no grounds at all. A conjecture may be "the formation or offering of an opinion on grounds insufficient to furnish proof", as the *Oxford New English Dictionary* records. That was what James Bernoulli's *Ars Conjectandi* was about.[1] But a guess is more typically a judgment of size, amount, number, distance, *etc.*, that has been formed at random or without any actual measurement or calculation. So it is hardly surprising that we tend to expect conjectures rather than guesses from those engaged in the professional pursuit of intellectual enquiry, and guesses rather than conjectures from those who are playing games.

The two terms are like one another, however, in being applicable not only to a certain kind of mental act but also, in a derivative sense, to the speech-acts which purport to declare the performance of such a mental act. For example, one can say not only (in the mental act sense)

> He long ago guessed that p, though he kept his guess to himself

but also (in the speech-act sense)

> In guessing as he did he pretended not to know the correct answer,

and similarly with the verb 'to conjecture'. I am concerned here mainly with the mental-act sense of these verbs, but now and again it will turn out important to notice that there is also a speech-act sense.

The same mental-speech act duality characterises the verb 'to speculate', at least where speculation is conceived of as an intellectual, not a financial occupation. One can say both

> He speculated for a long time about her motive though he kept his speculations to himself

and also

> In public he speculated about her motive though in fact he was privy to it.

But this is what we should expect, since speculating seems to be an activity constituted by occasional acts of guessing, conjecturing, or even fancying and imagining, along with associated meditations, question-posings, *etc.* A man's speculation about a topic can go on for years, and during that period he can make many distinct conjectures on different occasions. But what goes on for years, *stricto sensu*, is the activity of speculation, not the act of conjecturing. Hence one can say that he was speculating *about* her motive and guessing *at* it. Her motive formed the *focus* around which his activity of speculation centred. But the motive was thus also the *target* for each individual act of guessing. Consequently speculation runs a greater risk than conjecture of bringing intellectual disrepute on its author. One or two conjectures, clearly labelled as such, may stimulate interest, provoke curiosity and provide material for further research: conjectures invite proof or disproof. One can even begin one's thinking about a subject by speculating and end by having a good hypothesis to test. But universal and continuous speculation is an activity that never goes beyond conjecture. To describe a book or a theory as speculative is to imply that nothing but guesswork lies behind it. Hence while even the most reputation-conscious professors feel safe to hazard occasional conjectures, they may be less happy to be described as spending time on mere speculation. Speculation carries a whiff of intellectual disreputability because, being an activity and not just a single act, it may to that extent be taken to exclude, for some period, the occurrence of other intellectual procedures, such as the construction of proofs, the execution of experimental tests, and so on.

3. Guessing and suspecting

It may be tempting to think that a man's guesses and conjectures, like his suspicions, are what he is inclined to believe despite feeling that the evidence does not completely back him. But in fact guessing and conjecturing are rather different from suspecting.

One difference to notice is that suspicion constitutes a specific stage on a scale of conviction while guesses and conjectures are not tied to any particular point on this scale. Suspicion, though it may itself be assessed on a scale of strength and weakness, is readily contrasted with feeling sure and even with belief. Asked whether he believes that p, a man may reply

I'm not sure about it, though I do suspect it.

At least when he first has his suspicions he cannot feel sure about them. But guesses are different. A man may equally well distrust his guess or feel sure of it. Guesses may be either tentative or confident even though it is often rash, and sometimes just silly, to be confident about the correctness of a mere guess. So

John is merely guessing that p

is equally compatible with

John suspects that p,

with

John believes that p,

and with

John feels sure that p.

But it neither entails nor is entailed by any of these three propositions. Hence philosophers are not entitled to argue that so far as a hesitant examinee thinks he's guessing that p, he cannot believe that p. If there is any reason for supposing that he does not believe that p it must lie in the tentative, hesitating nature of his answer–in his obvious lack of conviction–not in the fact that he thinks he's guessing.

Another thing to notice is that suspicions are sometimes reasonable and sometimes unreasonable, but guesses can hardly be either. The question

Why do you suspect that p?

makes a good deal better sense than

Why do you guess that p?

If a wife has no good reasons for thinking that her husband has been unfaithful to her, but nevertheless suspects that he has been, her suspicion is unreasonable. But if in these circumstances she is asked to guess

whether or not he has been unfaithful and guesses that he has been, her guess is neither reasonable nor unreasonable. Because guessing is what we have to resort to when there is not enough evidence to draw conclusions from, guesses are not so readily classifiable into the sheep-and-goats categories of reasonable and unreasonable. They are all, in a way, non-reasonable, since they are a substitute for proper reasoning.

Accordingly, if a man replies to a request for information about the quickest way to reach central London

> I suspect that you should take the motorway,

he half endorses the truth of what he says. To that extent he risks his intellectual credit by saying it. If a lot of a man's suspicions turn out false we tend to conclude that he is a bad judge of evidence. But the man who says

> My guess is that the motorway will be quickest

does not commit his credit even to that extent. He has excluded any risk at all to his credit by making it quite explicit that he is guessing. If he is proved wrong, his reputation for knowledgeability, reasonableness or good judgement should not suffer, but only–perhaps–his reputation for being a good guesser.

We are perhaps more inclined to speak of a man's having made a reasonable, or unreasonable, conjecture that p. But this seems to be because, in one fairly strict sense of the word, as we have seen, a conjecture may need to be based on rather more grounds than a guess needs. Hence an assessment of reasonableness is not so out of place as it is for something described merely as a guess. Analogies, associations, *etc.* may hint that a conjecture along certain lines would not be far wrong, even if they do not constitute proof or confirmation of the hypothesis conjectured. And a conjecture that runs counter to current trends of expert opinion and flies in the face of plausible analogy might well seem rather irrational. Admittedly we do sometimes call something an educated guess, or informed guesswork. But this draws attention to the qualities or qualifications of the guesser rather than of the guess. Similarly when we call something a 'shrewd guess' we imply that it springs from the guesser's underlying experience and intelligence, not that it has any particular support from available evidence.

Correspondingly, where only guesses are available, the value of a guess, as a foundation for informed decision, has to be estimated from the past successes or failures of the guesser. If a man seems to be good at guessing the winners of horseraces, we may decide to place our bets in accordance with his predictions. Suspicions, however, *may* have reasons

behind them. So, other things being equal, they deserve to be taken more seriously than mere guesses. If I do have reasons for suspecting that Excalibur will win the 3.30, the reasons may be good or bad ones. Perhaps I have even heard a rumour that the other horses will be drugged. But if you know nothing about my past performance as a predicter and have no grounds for doubting my sincerity, you would clearly be more prudent to act on what I tell you about my suspicions than on what I tell you about my guesses.

Another point is that suspecting, like believing, is not much under our control. We can urge a man to have his suspicions, not request him. But we can request a man to make a guess about something, even in the mental-act sense–though of course we can't tell him what to guess.

4. Guessing, assuming and supposing

Since guessing thus differs from suspecting in excluding the consideration of reasons, it might seem rather like a kind of assuming (in the sense of making intellectual assumptions) or taking for granted. Where there is evidence to consider, it might seem, we form suspicions, draw conclusions or arrive at certainties. But where there is no evidence we have to fall back on assumptions–or guesswork.

This account would also seem to fit the fact that suspicion, unlike guesswork, involves a judgment of evidence, and is a stage on the way to certainty. People who announce their suspicions risk their intellectual credit to some extent, as we have seen, while people who announce their guesses do not. Now to declare that something is only an assumption, and not a claim or an assertion, is to declare that–for the moment, at least–one is not professing belief in it. By making such a declaration explicit one insures against loss of intellectual credit if the assumption subsequently turns out to be erroneous. So in this respect assumptions seem like guesses. Similarly both guesses and assumptions can subsequently be proved or disproved, verified or falsified, confirmed or undermined. Also both guesses and assumptions may later be shown by the facts to have been sensible, or foolish, in relation to one's own interests.

Thus there are substantial similarities between guessing and assuming, which lend some plausibility to the thesis that guessing is just a kind of assuming or taking for granted, or at least to the thesis that any proposition of the form

John guesses that p

necessarily implies a corresponding proposition of the form

John assumes that p.

But in fact there are substantial dissimilarities also between the two concepts. And it is arguable that, though the two actions may in practice sometimes coincide, overlap or provoke one another, yet guessing does not entail–*i.e.*, does not analytically imply – assuming.

A principle difference between describing something as a guess and describing it as an assumption is that the former imputes an interest in some question to which the guess itself constitutes an answer while the latter suggests that the direction of interest–*i.e.*, of the assumer's interest–is towards some other issue or project. A man guesses the result of the next cointoss, perhaps, and in doing so he may assume that the coin is not biased. But he wins the game if and only if his guess is correct, irrespectively of the truth of his assumption. There are guessing-games, but not assuming-games. So it's hardly surprising that, though guessing that p, as we've seen, sometimes analytically implies that p, assuming never does. One can guess a riddle, but not assume it. Similarly, guessing who was at last night's party suggests undisguised curiosity to know who was there: assumptions about the matter do not directly do so. Of course, an assumption is sometimes announced like the flying of a diplomatic kite, in order to ferret out information as to its truth, and so, when uttered with a rising intonation, sentences like

I assume that John was there

have come to constitute a polite form of interrogative speech. But the subtlety and politeness of this mode of enquiry depend on the fact that it looks the other way: the man who is described as assuming that p is imputed another direction of interest than towards the question whether p.

Again, a man may feel pretty convinced that it is going to rain, but decide to assume that it will not and leave his umbrella at home. In these circumstances he could not guess that it won't rain. If he's tapped the barometer and looked at the sky and concluded that rain is likely, he's not made any kind of intellectual error if he decides to assume nevertheless that it will be fine. Indeed if he tells us only that this is the assumption he's making, and tells us nothing about his conclusions from the barometer and the clouds, he's not necessarily misleading us in any way. But if he were to tell us in those circumstances that his guess is that it will be fine, he would be misleading us, since if he were in fact reduced to guessing he cannot have had that amount of evidence, let alone concluded the opposite.

It is tempting to say too that assuming is rather more under our control than guessing is, and even more so than suspecting. But if such a thesis is to be maintained it must be qualified in certain respects.

First, we can normally decide whether or not to speculate about a given topic, if we cannot draw any proper conclusions about it. So if there is something about the mental act of guessing that is not fully under our control, it pertains more to the content of the guess made than to the making of it. We cannot choose our guesses, though we can choose to guess now, say, rather than to wait and see.

Secondly, though we are free to assume what we like–and people sometimes even make self-contradictory assumptions–our assumptions are sometimes implicit, tacit or unconscious. So, since not all assumptions are made in the full awareness that they are being made, not all assumptions that p are the outcome of decisions expressly to assume that p. But when someone draws our attention to the fact that we have been tacitly assuming that p we can in principle always decide whether we shall continue to assume it in this way or prefer rather to reject whatever theories or policies committed us implicitly to that assumption.

But, subject to the above qualifications, it is clear that a man's assuming is more under his voluntary control than is his guessing. Indeed, even when a man has in fact guessed that p he may prudently decide to assume that not-p. He may guess, for example, that the stock-market will remain bullish over the next week but decide not to assume it, since he does not want to risk the heavy financial loss he would suffer if his guess were wrong. So he converts his stock-market holdings into liquid assets and thus lets his assumption run counter to his guess. (I haven't time to discuss presuming properly. We tend to treat it rather as a special form of assuming–*viz.*, assuming in the face of possible objections–and as such it has all the differences from guessing that assuming has.)

In the face of all this it seems quite impossible to accept the thesis that guessing is just a form of assuming or that statements about guesses necessarily imply corresponding statements about assumptions. Nor is guessing much closer to supposing than to assuming.

In one sense of the verb, to suppose that p is to consider the possibility that p, as distinct from taking it for granted, which is what we do when we assume it. Assumptions provide a basis for proofs or arguments, while suppositions open up avenues of enquiry or investigation. Accordingly supposing, in this sense, requires one to put oneself into an appropriate frame of mind, and it may therefore sometimes be a little more difficult than assuming. It would be easy enough to suppose here in Bloomsbury, that the level of water in the Thames had risen two inches since yesterday, and rather more difficult to suppose–seriously–that it had risen two hundred feet. But even the more difficult supposition could be carried through, if one steeled oneself to it. So in principle,

at least, one can suppose (in this sense), as well as assume, that not-p while guessing that p.

In another sense of the verb to 'suppose', supposing is a form of judgment in which we indulge when knowledge is not to be had. In this sense one can say

> She supposed that he had missed his train and therefore waited another half hour.

But in this sense suppositions resemble suspicions, in that there may well be reasons for them, unlike for guesses, and to that extent a man who announces his supposition puts his intellectual credit at risk. Also supposing, in this sense, has an intermediate position on the scale of conviction that runs from suspecting to feeling sure, while guessing is compatible with different degrees of conviction. So guessing is at least as different from supposing, in this sense, as it is from suspecting.

5. Guessing and jumping to a conclusion

If guessing is not a form of assuming or supposing, perhaps instead it is a degenerate form of inference or, if you prefer it, an alternative to inference? If one cannot guess, though one can assume, what one likes, then at least in this respect guessing is like inferring. For, however much one may want a certain conclusion to emerge, one cannot draw it in one's mind unless one thinks it actually follows. Wishful thinking may occur, but when it occurs the thinker is a victim, not an agent, of it. Note too how 'conclude', 'infer', 'estimate', 'calculate', *etc.* have speech-act as well as mental-act senses, just like 'guess', 'conjecture' and 'speculate'. And the fact that the speech-acts can be performed at will does not imply that the corresponding mental acts can. So perhaps guessing may be defined as that form of inference which reaches a more positive conclusion than the guesser's evidence justifies?

A briefly stated analysis along these lines was put forward by Robert Fogelin (in *Evidence and Meaning*, 1967, p. 73). His thesis was that

> John guessed that p

meant the same as

> John adopted p and John lacked adequate grounds for p.

Fogelin's analysis suffers from the disadvantage that adopting or not adopting a proposition, though its precise nature may be a matter of dispute, certainly looks like something subject to control by voluntary decision. But the analysis might nevertheless be right to require that

when John guesses that p he must lack adequate grounds for thinking that p. Certainly when a hesitant examinee believes that he is only guessing it is because he lacks the evidence of conscious memory.

This theory has the great advantage that it integrates the concept of guessing into a familiar semantic field–a familiar region of the language-map. Evidential adequacy is certainly a topic that logicians and epistemologists have often explored. But the theory will not do as it stands because it fails to distinguish guessing either from conjecturing or from jumping to conclusions. One can lack adequate grounds for thinking that p both if there aren't enough objectively relevant data to justify any kind of conclusion and also if there are enough objectively relevant data to justify some kind of conclusion but the data give merely partial support to the conclusion that p. In general guessing seems to belong in the former context, where the guesser is either very short of any objectively relevant data at all or has not paid adequate attention to the data that are available, while conjecturing or jumping to conclusions is what occurs in the latter context. A man conjectures, in the strict sense, when he himself thinks that the evidence gives only a partial support to his conclusion. He jumps to a conclusion, when he substantially overestimates the extent of the support provided by the evidence for the conclusion he draws. Guessing is done when there is little or nothing to go on, but jumping requires something to jump from. Guessing is a form of inference that is degenerate through paucity of evidence, not one that is invalid through incompleteness of support.

For example, you guess the weight of a parcel by holding it in your hands and using no balance or scales to measure it. But you jump to conclusions about its weight if you put it on the scales and then assume, without evidence, that the scales are marked off in pounds rather than in kilograms. Again, suppose that there is a lot of evidence about a certain crime and on the balance of probabilities it seems likely that Jekyll is the criminal, though his guilt cannot be said to be beyond reasonable doubt. Would a juror who concluded that Jekyll was *undoubtedly* guilty be correctly described as guessing this? One is inclined to say rather that, though there may well have been some element of guesswork involved in the juror's coming to his conclusion (perhaps he guessed that the alibi testimony was perjured), nevertheless the overall conclusion is more correctly described as a rash or over-positive one than as a guess. Now contrast the case where a man is told that a penny will be tossed twice and unhesitatingly predicts that it will fall heads uppermost at least once. Clearly this conclusion also has the balance of probabilities in its favour, since the coin falls heads uppermost at least once in three out of the four possible ways in which it can make its pair of falls, and

the conclusion is also one that is not fully supported by the data. But here the prediction would normally be described as a guess, rather than as a case of jumping to conclusions. For, though the prediction has a relatively high (.75) mathematical probability of being correct, there is no *evidence* to support that particular prediction.

But, though this is in general the right way to distinguish between guessing and jumping to a conclusion, one or two qualifications need to be added. A man might be told the barometric pressure, wind speed, cloud height, *etc.* on a given occasion and still be correctly described as merely guessing that it will rain, because he does not know that these data are relevant evidence for meteorological predictions. Or he might know *that* they are relevant, but not know *how*, because he does not know the content of any laws connecting these present phenomena with future ones; and if he predicted rain in such circumstances he would still be guessing. Or the man might know the data and the laws but not try to work out what their combination implied; and again his prediction would be a mere guess. So it is not paucity of evidential data as such that matters, but rather paucity of evidential data that have been taken into account by the guesser. Analogously, if a man knows the axioms of a deductive system and its rules of proof, while not thinking that he knows, or knows of, any proof that p, his coming to think it provable that p would be a mere guess or conjecture. But if he mistakenly thinks that he has now found such a proof, he would be jumping to a conclusion.

Also it needs to be borne in mind that, for a man to be truly said to be guessing, his present performance must not be the result of any even quite distantly previous taking into account of relevant data. If he long ago learned on good authority that Elizabeth Tudor died in 1603 and as a *result* gives this date when now asked, then even if he himself is unaware of the connexion between his previous learning and his present performance it would nevertheless be incorrect to call the latter a 'guess'. "He may think he's guessing", we should say, if we were sure of the connexion, "but in fact he is not."

6. Guessing and hypothesising

So much for what might conveniently be termed the upper threshold of guessing–the point beyond which guessing turns into estimating (however rashly) or inferring (however overconfidently) or into genuine knowledge. Now what about the possibility of a lower threshold? Are there any limits to the wildness of a guess? Are there limits beyond which we should not say that a man was guessing but that he was do-

ing something intellectually inferior even to guessing? Is hypothesising perhaps something of this nature?

Obviously it would be odd if I ask "Who do you guess will succeed Heath as British Prime Minister?" and you answer "Bob Hope". We are all aware that the well-known Bob Hope is American and not British; a comedian and not a politician, and so on. So, unless you are referring to someone else of the same name, or are referring to someone else and have got his name wrong, you can't be guessing seriously. But what is flippant or non-serious here is a speech-act, not a mental operation. Indeed in that (speech-act) sense of the verb 'to guess' there is something wrong about a guess that p–it is flippant, dishonest, misleading, or what have you–whenever the speaker does not also guess that p in the mental-act sense of the verb. The man who speech-act guesses that the next Prime Minister of the United Kingdom will be Bob Hope is pretending to have mental-act-guessed that this will be so, just as the man who speech-act-concludes this would be pretending to have drawn that conclusion in his own mind.

But in the mental-act sense of 'guess' there cannot be any flippant, dishonest or misleading guesses, unless the possibility of self-deception is admitted in such cases. Perhaps a man who was twisted up with worries and fears about his symptoms, for example, might pretend to himself that he was only guessing at the nature of his illness. But a case of this type would be relevant to the upper threshold of guessing, not the lower. It would be an instance of inference or suspicion masquerading as guesswork. One finds it much more difficult to imagine what it might be like for something that's not even a guess to masquerade as one in a man's mind. How could I pretend to myself that I was guessing that Bob Hope would succeed Heath when in fact I was not in the slightest bit inclined to think this? Of course, some guesses are rather wild ones, in that they have not been curbed by any constraining factors in the situation. If I really did guess that Bob Hope would succeed Heath, I should be guessing very wildly indeed, and if someone looked at me and guessed that I was just twenty-one he would be making a very stupid guess at my age. Most respectable guesses are made within a range of possibilities that has certain obvious limits. But awareness of these limits does not count as knowledge of relevant evidence because it does not substantially favour one possibility within the limits over another. Also, guesses can be very tentative or provisional if the guesser feels he may want to change his mind, and they can be very poor if they come nowhere near their target.

Nevertheless, though there are certainly various dimensions of appraisal for guesses, none of them seems to direct our attention at its

extremes to mental operations which are not even guesses. The merest of mere guesses has no data at all to go on, or takes no account at all of the relevant data that are available, and that is about as low as you can get in ways of thinking up solutions to an intellectual problem.

How then do guessing and conjecturing relate to hypothesising? If there is no scale on which hypothesising can be located somewhere below guesswork is forming a hypothesis actually a kind of guesswork?

To refer to a proposition as Professor Smith's hypothesis is to imply that at some time Professor Smith actively entertained the possibility that it constituted a solution of some problem in which he was interested. So hypotheses, like guesses and conjectures, have authors and dates at which they first come into their authors' minds. Also, like guesses and conjectures, they can later be discarded or modified, because the evidence turns out to be wholly or partly against them. Again, to call someone's thought a mere hypothesis, like calling it a mere guess, is to imply that its author has not arrived at it by making a valid inference from the evidence. Finally, in at least one respect hypothesising is more like guessing than it is like assuming. For, as we have already seen, assuming differs from guessing in that a man's assumption that p is normally in aid of the solution of some other problem than the question whether p, while his guess that p formulates an answer to just that question. And correspondingly we normally form the hypothesis that p when we are interested in a question to which the hypothesis itself affords an answer, as in historical research or scientific enquiry. (I disregard here the use of the word 'hypothesis' in the language of mathematical proof, where the phrase 'proof on the hypothesis that p' does not differ in function from the phrase 'proof on the assumption that p'.)

These similarities between guessing and hypothesising have no doubt lent some plausibility to the thesis that guesswork or conjecture is an essential element in scientific enquiry.[2] However, that thesis appears substantially less plausible when we consider the important differences that also exist between guessing and hypothesising.

The first obvious difference is that guesses and conjectures can be communicated as assertions (in the ordinary, nonphilosopher's sense of 'assertion') while hypotheses cannot. The following piece of narrative, for example, contains no logical or linguistic oddity:

> 'Excalibur will win the 3.30', he asserted. 'That's only a guess', she commented. 'Quite right', he replied.

But to communicate an hypothesis as such, a man would have to say, not

> Excalibur will win the 3.30

but either

My hypothesis is that Excalibur will win the 3.30

or some sentence like

Perhaps Excalibur will win the 3.30.

Admittedly a bookmaker could perfectly well assert that Excalibur will win the 3.30 when really this is only an hypothesis of his which he is testing out on his audience by voicing an assertion that challenges refutation. But then his utterance would either not communicate that he has this hypothesis or would only communicate that fact in virtue of being also a way of gathering evidence for or against the hypothesis. The underlying difference here is a difference of commitment. The man who guesses or conjectures that p is committed in his own mind, albeit minimally and temporarily, to the proposition that p. Though he may not actually believe that p, he must at least be inclined to believe that p. That is why the announcement of his guess can take the form of an assertion. But the man who merely hypothesises that p is carefully refraining from committing himself to this proposition. He commits himself rather to the proposition that it is worth investigating whether p. I do not mean that if a man guesses that p he cannot also be hypothesising that p. He may be. But the element of commitment that is ascribed when we talk of him as guessing is not ascribed when we talk of him as hypothesising. Hence if we say that he is *merely* hypothesising we imply the actual absence of this element. Correspondingly, if a man knows that the proposition that p is inconsistent with the proposition that q, then he cannot guess or conjecture at the same time both that p and that q, just as he cannot properly assert both propositions. But so non-committal is the adoption of an hypothesis that at one and the same time he could quite properly be pursuing some investigations on the hypothesis that p and others on the hypothesis that q. For example, if there are two rival suspects in a murder case the detective who orders both their houses to be searched might be described as investigating two mutually inconsistent hypotheses.

Secondly, guessing is a mental operation that does not need to carry with it any intentions about future action. But hypothesising is not such an operation. In addition to formulating the proposition hypothesised one must also intend to act on it for a while, or to check it, or at least intend someone to consider it somehow. Etymologically it's the putting of one thing under another, and so the laying of a basis on which to proceed. Hypothesising is therefore a forward-looking operation in a way that guessing is not. Admittedly, conjecturing has a more forward-looking aspect than has guessing, as we have seen. To describe someone

as conjecturing that p seems normally to imply that he is raising, or at least airing, the question whether p. But conjecturing still does not connote any intention on the part of the conjecturer. A conjecture may be left to lie on the record, whereas the hypothesiser characteristically intends something to be done.

Thirdly, hypotheses are invented and patronised by their authors, while guesses and conjectures just occur to them. So in at least one respect the mental act of hypothesising is more like that of assuming than like that of guessing. It is more under a man's voluntary control than his guessing is. A man can choose his hypotheses, as he can choose his assumptions, but – as we have already seen (§4)–he cannot choose his guesses.

Fourthly, to talk of someone's hypothesis is to imply nothing about how he reached his state of mind. But to talk of his guess does have such an implication. It implies that he did not reach his state of mind by inference from given premises. Indeed, a man may examine quite a lot of evidence before he picks on a particular hypothesis as being the one that would provide the best explanation of known phenomena if it survived testing, or as promising the most rewarding results if further investigated. But in these circumstances his adoption of the hypothesis obviously cannot strictly be described as a guess though it might well be described as a conjecture, since he does not suffer from paucity of the kind of evidence that may reasonably be expected to be available at this stage. Admittedly, if he does start from a serious paucity of evidence, he can both guess that p and also decide to treat that proposition as an hypothesis to be tested. But this would be a contingent identity, not a necessary one.

Finally, it is quite possible for a man to both hypothesise that p and guess (or conjecture) that not-p, just as he can also, as we have seen, assume that p while guessing (or conjecturing) that not-p. A familiar form of this type of situation occurs when a man feels intellectually bound to investigate a plausible hypothesis even though he has a strong hunch that it is wrong. There is nothing at all absurd or self-contradictory, whether logically or pragmatically, in a pathologist's saying

> My current working hypothesis is that all cancer is caused by viruses, but my guess (or conjecture) is that I shall be able to refute this hypothesis within the next five years.

In short, hypothesising neither entails nor is entailed by guessing. So guessing, conjecture and speculation have no intrinsic role to play in scientific method unless some form of guessing or conjecture, though not analytically necessary for arriving at testable hypotheses, can nev-

ertheless be shown to be the professionally correct, or intrinsically most suitable, way to arrive at them. But there are at least two good reasons why neither guessing nor conjecture should be so regarded.

First, except in the case of very novel and pioneering enquiries, there is normally at least some relevant evidence available from the results of previous investigations in the same field, and it would obviously be foolish and irresponsible not to pay due attention to that evidence. But in very many cases this excludes the possibility of mere guesswork in the formulation of hypotheses, though it is compatible with the possibility of conjecture. So a scientist who hypothesises that p may do so because he guesses or conjectures that p, or because he suspects that p (and I have already discussed the difference between guessing and suspecting), or because he guesses, conjectures or suspects that not-p and wants to clear the hypothesis that p out of the way, or because he has some other reason for thinking it worth while to test the hypothesis, such as the prospect of coming across some interesting facts in the course of his tests. Though brilliant guesses or conjectures are undoubtedly impressive, they are not the life-blood of science.

Secondly, making a guess or conjecture that p, instead of just adopting the hypothesis that p, entails a slight, but often undesirable, diminution in the suspense of judgment that professional prudence dictates at the stage at which a hypothesis has not yet been tested. Hypothesising that p, as we've seen, is compatible with total suspense of commitment on the question whether or not it is true that p. But guessing and conjecturing, like sincere assertion, is not compatible with this kind of total non-commitment. No doubt a certain degree of commitment to the truth of an hypothesis helps some people to have the determination to persist in the arduous task of testing it. But such a commitment may also blind them to the worthwhileness of concurrently investigating other hypotheses, or even unconsciously deter them from testing their own hypothesis as severely as possible.

Moreover the main objectives that certain philosophers have tried to further, by claiming that science essentially involves guesswork or conjecture, are not in fact furthered by such a claim (whether or not these objectives are themselves correct). These objectives seem to be either (1) to emphasise the inevitable element of uncertainty in experimental findings (Carnap) or (2) to argue that hypothesis-formation precedes, or should precede, the search for experimental evidence (Popper). Let us consider these two objectives separately, and examine each in relation to the roles of both guesswork (a) and conjecture (b).

1a. To feel that the evidence supports a particular hypothesis but does not give it absolute certainty, is by no means the same as guessing the truth of the hypothesis, since guessing takes place where there is no evidence at all. If all science were guesswork, Carnapian inductive logic could never have any application to it.

1b. Nor should the feeling (that all science involves uncertainty) be identified with conjecturing the truth of the hypothesis, since though conjectures are based on evidential indications and so have to be formed after some such indications have come to one's notice, the feeling that the evidence supports a particular hypothesis may relate partly to evidence that one has noticed after initially conjecturing the truth of the hypothesis. A confirmed conjecture is not a mere conjecture.

2a. Even if an hypothesis is, or should be, formed before experimental evidence is sought, it does not follow that the hypothesis that p is reached by guessing that p. As we have seen, a man may quite well hypothesise that p and guess that not-p.

2b. Still less does it follow that the hypothesis is reached by conjecture. Conjectures in the strict sense, as we have seen, are always based on prior attention to supposed evidential indications.

No doubt some defenders of these philosophers will argue that their use of words like 'guess' or 'conjecture' has always been a loose one, and that its sense is to be understood from the contexts in which it is embedded. If so, I am content to have my own remarks interpreted as arguing that what these philosophers have said when they used these words was misleading rather than mistaken.

Notes

1. Cf. "Conicere rem aliquam est metiri illius probabilitatem", *Ars Conjectandi*, IV, ii.

2. For those whose native language is German the thesis has perhaps an additional source of plausibility, since the same word raten means both to guess and to counsel and so some of the professional expertise associated with counselling may rub off on to the meaning of raten when a philosopher like Popper writes: "Wir wissen nicht, sondern wir raten" (*Logik der Forschung*, 1935, p.207).

13

BAYESIANISM VERSUS BACONIANISM IN THE EVALUATION OF MEDICAL DIAGNOSES

1.

Medicine is an excellent example of a field in which it is often necessary to make inferences under conditions of uncertainty, either because further delays in therapy would be too risky or because further tests or examinations would be too costly. In the words of Sardeman, 1964, p. 142, 'If you did all laboratory tests to everybody you would never get finished, and nobody would be able to pay for it.' But some uncertain inferences are less reliable than others, and it is important to be able to draw the right distinctions here. So how is diagnostic uncertainty to be graded? Two models for the evaluation of medical diagnoses are often found in apparent conflict with one another today. The first is avowedly Bayesian. The second is demonstrably Baconian in its logic though not normally conceiving of itself as such. The purpose of the present paper is to clarify the nature of the difference between these two models and to show that, when restricted to their legitimate roles, they do not conflict with one another.

Some initial clarification of terminology may be helpful. By 'evidence' I shall mean one or more facts of the following four kinds:

1 facts about the patient's social or physical environment, or his occupation, or life-style, or biography,

2 symptoms, *i.e.* the patient's complaints,

3 signs, *i.e.* objectively observable manifestations of the patient's state of health that the physician himself can detect, and

4 the results of laboratory tests, of roentgenography or of any other process that generates data from the patient's body with the aid of paraclinical procedures.

(All four kinds of evidential facts may be heavily 'theory-laden', in the sense that their present recognition in particular cases is the result of previous scientific effort in regard to classes of similar cases.) By a 'disease' I shall mean any causal process within a person's body whereby one or more symptoms and/or signs of ill health is produced. Of course, many such causal processes are initiated from without–by a blow on the head, say, or by the invasion of an infecting parasite, or by contact with a carcinogenic substance, *etc.* But I am confining my attention here to diagnoses that are instrumental for therapy, as distinct from those appropriate to preventive medicine. So it is convenient to restrict the concept of diseases here to causal processes within the body. Again, there are some commonly associated groups of symptoms and/or signs of ill health for which no satisfactory aetiology is yet available. I shall refer to them as 'syndromes', and assume that any report of them in a particular patient is a description, not a diagnosis.

In these terms the progress of pathology consists in replacing the description of syndromes by the diagnosis of diseases. But this process of replacement is an open-ended one, as the chain of causation is traced backwards. Modern nosology begins in the seventeenth century with Sydenham's conception of syndromes as natural species; it makes a great leap forward in the nineteenth century when the practice of autopsy begins to reveal macro-anatomic causes of syndromes and thus confirms the scientific respectability of conceiving a disease as a causal process; it makes another great leap forward in the following decades when the vast proliferation of paraclinical tests makes it possible to trace many disease-processes further back and permits the classification of many diseases in terms of microentities or microstructures that are studied within newly specialised disciplines (bacteriology, virology, haematology, immunology, *etc.*); and one can see that yet another great leap forward is conceivable if many more diseases (beyond the familiar industrial and nutritional ones) could be traced back to the social or environmental factors that foster or provoke them. A nosology is thus conceivable in which, for the purposes of preventive medicine, it would be convenient to conceive of all disease as a causal process that includes social and environmental events as well as bodily ones, and it would then be right to modify the classification of diseases accordingly. But in the present state of knowledge, and

for the purposes of most curative medicine, a more restricted conception of disease is appropriate. Even then the causal end-point of diagnostic analysis and nosological definition is often not determined uniquely. To cite an example given by Feinstein, 1974, p. 217, is myocardial infarction a satisfactory end-point for an appropriate type of chest-pain, or should the series of admissible diagnoses extend onward from myocardial infarction to coronary thrombosis, coronary atherosclerosis, hyperlipemia, nutritional imbalance and neurotic personality?

2.

The Bayesian theory appears in various forms. What is essential to it is that the validity of a diagnosis is thought of as a conditional probability –the probability that the patient has a certain specified disease, given certain specified evidence–and that this conditional probability is taken to be a function partly of the probability of the evidence, given that he has the disease, and partly of relevant prior probabilities. In its simplest and most straightforward form the theory is just exploiting the familiar principle in the mathematical calculus of chance

$$p(D/E) = \frac{p(E/D) \ \times \ p(D)}{p(E)}$$

where $p(E) > 0$, with 'D' standing for the name of the diagnosed disease and 'E' for the record of the evidence.

But as some Bayesian writers on medical diagnosis, e.g. Ledley and Lusted, 1959, point out it is possible to eliminate $p(E)$ here, so long as it is safe to assume that in the particular case under examination E is caused by just one of the diseases D_1, D_2, ..., D_n. If D_k is now the diagnosed disease we can write

$$p(D_k/E) = \frac{p(E/D_k) \ \times \ p(D_k)}{\sum_{i=1}^{n} p(E/D_i) \times p(D_i)}$$

An arguable merit of the theory is thus that it directs a clinician's attention to the importance of taking into account the value of $p(D)$ that is appropriate for his own patients: some kinds of disease may be much commoner among those already admitted to a hospital, for instance, than among those who consult their general practitioner, as Wulff, 1976, pp. 83-4 points out.

A number of objections have been made to the proposal of Bayesian evaluations for medical diagnoses, *e.g.* Feinstein, 1974, 276 ff., Lange, 1976, p. 16, Jacquez and Norusis, 1976, p. 381, and Feinstein, 1977. These objections fall into four main categories.

The first type of objection argues that a Bayesian account seeks merely to determine the probability with which a certain nosological term fits a given array of assembled data. Bayesians are therefore accused of being blind both to the importance of understanding the causal processes that have generated the data and also to the sequential nature of diagnostic procedures and their frequent incorporation of feedback, whereby initial examinations suggest tests, test-results suggest therapy, a patient's positive or negative response to therapy may suggest further lines of investigation, and so on. The situation to be represented is very much more complex than the Bayesian formula is capable of representing, and in any case clinical reasoning often terminates in therapeutic action rather than diagnostic conclusion.

But the Bayesian has a ready answer to at least part of this objection. He merely points out that 'D' in '$p(D/E)$' can stand for any kind of event, property or process that may be inferred, at the appropriate level of probability, from E, where E is the evidence available at any one stage of the diagnostic investigation. So a sequential form of Bayes' theorem can represent the sequential nature of diagnostic procedures. Obviously a Bayesian account could have no right to claim that it revealed heuristic strategies which are specific to the medical situation, such as the investigation of underlying causal processes or the interspersion of diagnosis with therapy. But that would not debar it from claiming an ability to grade the probability of the inference in which any such strategy issues. Moreover, a Bayesian account here is on much safer ground if offered as an idealised reconstruction of the relevant clinical evaluations rather than as a practically applicable methodology for making them, though in this relatively unambitious form it may be of greater interest to philosophers than to physicians.

A second type of objection stresses the risks incurred in assuming that in the particular case under examination the pathological manifestations result from just one of the diseases $D_1, D_2, \ldots, D_n$. Suppose yet another disease could cause such manifestations, or suppose that in the particular case they were caused by the presence of two or more of $D_1, D_2, \ldots, D_n$?

But the Bayesian might legitimately reply that we can only take into account what we already know. If additional disease-candidates become available we shall naturally have to revise our estimates. Similarly if two or more of $D_1, D_2, \ldots, D_n$ can co-occur, the expression that may replace $p(E)$ in Bayes' law becomes correspondingly more complicated. We have to consider not $\sum_{i=1}^{n} p(E/D_i) \times p(D_i)$ but the sum of

$$p(D_1 \& \overline{D_2 \vee D_3 \vee \ldots \vee D_n/E}) \times p(D_1 \& \overline{D_2 \vee D_3 \vee \ldots \vee D_n})$$

and

$$p(D_1\&D_2\&\overline{D_3 \vee D_4 \vee \ldots \vee D_n/E}) \times p(D_1\&D_2\&\overline{D_3 \vee D_4 \vee \ldots \vee D_n})$$

and so on.

A third type of objection argues that a Bayesian account is only able to accommodate the corroborative force of further evidence by assuming that each evidential fact is independent of every other, whereas such independence is rarely, if ever, to be found in diagnostic data. This objection is not normally worked out in detail. But if expressed in exact terms it would presumably run somewhat as follows. Or at any rate the objection seems to lack plausibility unless this is what it means.

An instance of corroboration may be represented by the inequality $p(D/E_l 1\&E_2) > p(D/E_1)$. Here E_2 corroborates E_1's support for D. In Bayesian terms we therefore need to seek the conditions under which

$$\frac{p(E_1\&E_2/D) \times p(D)}{p(E_1\&E_2)} > \frac{p(E_1/D) \times p(D)}{p(E_1)}$$

These conditions will obviously be the same as those under which

$$\frac{p(E_1\&E_2/D)}{p(E_1\&E_2)} > \frac{p(E_1\&E_2)}{p(E_1)} \tag{13.1}$$

Now on a reasonable assumption that $p(E_2/D) > p(E_2)$, *i.e.* that the relevant sign, symptom, *etc.* is commoner among those with the disease in question than in the population at large, we can be sure that

$$\frac{p(E_2/D)}{p(E_2)} > 1$$

Correspondingly we can be sure that

$$\frac{p(E_1/D) \times p(E_2/D)}{p(E_1) \times p(E_2)} > \frac{p(E_1/D)}{p(E_1)} \tag{13.2}$$

So if we identify conditions under which (13.1) is deducible from (13.2), we thereby identify conditions under which E_2 corroborates E_1 in supporting the inference to D. Now an obvious condition under which (13.1) is deducible from (13.2) is when E_l and E_2 are independent of one another in the sense of being neither positively nor negatively relevant to one another, *i.e.* $p(E_2/E_1) = p(E_2)$ and $p(E_2/E_l\&D) = p(E_2/D)$. But is it plausible to suppose that this is at all commonly the case? Compare Jacquez and Norusis, 1976. To the extent that people don't have any symptoms or signs of ill health unless they have some disease we should

certainly expect that, when E_1 and E_2 are capable of co-occurrence as symptoms or signs of the same disease, $p(E_2/E_1) > p(E_2)$. The very existence of a disease, D, that can produce both E_l and E_2 is bound to generate this inequality. Therefore, the objection would run, we should not normally expect to find the kind of independence under which, according to the Bayesian theory–e.g. Lusted, 1968, p. 17, cf. Tversky and Kahneman, 1977, p. 189 –corroboration takes place. Yet if that theory can give no account of corroboration it is worthless as a reconstruction of the reasoning operative in medical diagnosis, since it is a universally accepted feature of diagnosis that a suitable combination of supporting evidential facts gives added strength to the conclusion.

But the objection is less powerful than it seems. It rests on supposing that the specified independence condition is necessary, as well as sufficient, for corroboration to take place, and in fact this is not so. There are other, more appropriate conditions which are also sufficient to entail corroboration. Though Bayesian calculations do not normally rely on the existence of these conditions, the Bayesian account can be rescued by recognising their significance. Specifically, it can be proved that $p(D/E_1\&E_2) > p(D/E_1)$ if the following six conditions hold good:

$$p(D/E_1) > p(D) \tag{13.3}$$

$$p(D/E_2) > p(D) \tag{13.4}$$

$$p(E_1\&E_2) > 0 \tag{13.5}$$

$$1 > p(D/E_1) \tag{13.6}$$

$$p(E_2/D) \leq p(E_2/E_1 \ \& \ D) \tag{13.7}$$

$$p(E_2/D) \geq p(E_2/E_1 \ \& \ D) \tag{13.8}$$

A formal proof is given in Cohen [1977].

Now of these six conditions (13.3)-(13.8) the first four would in any case be present in any normal diagnostic situation where corroboration takes place, and a moment's reflection should suffice for anyone to see that the last two conditions are scarcely an unreasonable requirement. What (13.7) asserts is that, given the *presence* of the disease, the existence of the symptom, sign or other evidential fact E_2 is not rendered

less probable by the existence of E_1. After all, in a disease where two symptoms characteristically exclude one another, their co-occurrence can hardly *reinforce* diagnosis of the disease in question. What (13.8) asserts is that, given the *absence* of the disease, the existence of E_2 is not rendered *more* probable by the existence of E_1. After all, where one symptom is particularly inclined to accompany the other when the disease is *absent*, co-occurrence of the two symptoms can again hardly *reinforce* diagnosis of the disease in question.

In fact these two inequality conditions are precisely analogous to those under which the testimony of one witness corroborates that of another, or two pieces of circumstantial evidence converge to raise the probability of a suspect's guilt: cf. Cohen, 1977, p. 101 ff. Here 'D' can be interpreted as the point at issue, 'E_1' as the first witness's assertion of this point, and 'E_2' as the second witness's assertion of it. One witness should not be more inclined to contradict the other on those occasions when the latter's testimony happens to be true, since agreement of the two testimonies would then tend to indicate their incorrectness. So, if one witness is to corroborate another, the former's inclination to give true testimony must not be reduced, but may be increased, when the latter's testimony is true. Equally, one witness must not be more inclined to give false testimony about the point at issue when the other witness does so: for example, there must be no conspiracy to deceive. Thus, both (13.7) and (13.8) express conditions that would properly be required to hold if the testimony of two witnesses to the truth of the same proposition is to be better than that of just one. On the other hand, if E_2 had to be independent of E_1, in the sense of E_1's being neither positively nor negatively relevant to E_2, many apparent cases of testimonial corroboration could not count as such. For example, since most people are honest most of the time $p(E_2/E_1)$ is surely greater than $p(E_2)$ in most cases. Hence *this* kind of independence is just as inappropriate a requirement for testimonial as for diagnostic corroboration. The formalism is right not to include it.

Similar considerations apply also to the convergence of circumstantial evidence, as when it is shown that a murder-suspect had both motive and opportunity. But if one looks for a situation in which an independence condition of this kind *does* underwrite corroboration–*i.e.* a situation where $p(E_2/E_1) = p(E_2)$ and $p(E_1/E_1\&D) = p(E_2/D)$–one can apparently find it sometimes in games of chance. For example, let 'D' be interpreted as 'the die falls 6 uppermost'. 'E_1' as 'the die falls with a number > 2 uppermost' and 'E_2' as' the die falls with an even number uppermost'. Here $p(E_2/E_l) = p(E_2) = \frac{1}{2}$, $p(E_2/E_1\&D) = p(E_2/D) = 1$, $p(D, E_l\&E_2) = \frac{1}{2}$ and $p(D/E_1) = 1$. The inference to be drawn is that,

pace Hume [1739] book I, part III, section XII, and many others, games of chance are a rather poor model for probabilistic reasoning about human conduct and physical causation. Corroboration, or evidential convergence, has an importantly different structure in regard to games of chance from what it has in regard to empirically based probability judgments.

Of course, what emerges here is an analysis of corroboration that is employable by any account of diagnostic reasoning in terms of mathematical probabilities. It is not a specifically Bayesian analysis. But it is quite consistent with Bayesian principles, and is a satisfactory rejoinder to the objection about lack of independence between symptoms, signs, *etc.* E_2 will not normally corroborate E_1 at all unless conditions (13.7) and (13.8) are satisfied alongside (13.3)-(13.6), whatever the nature of the calculations that are used to evaluate $p(D/E_1\&E_2)$. Moreover, a Bayesian analysis does help to bring out the nature of the conditions that affect the degree of corroboration which takes place. The analysis shows that the degree of corroboration in fact increases as

$$\frac{p(E_2/E_1\&D)}{p(E_2/E_1)} \tag{13.9}$$

increases. The rationale of this lies in a medically familiar fact: *cf., e.g.* Jacquez [1964], p. 31 f. We have to distinguish genuinely additional evidence from mere redundancy, wherever disease-manifestations tend to occur together, depending on the nature of the causal linkage. Where the manifestations are closely connected together, irrespective of whether or not the disease itself is present, the level of (13.9) will be low, because $p(E_2/E_1)$ will be high relative to $p(E_2/E_1\&D)$, and corroboration will be correspondingly slight. But, where the manifestations are connected together only through the causal mechanism of the disease itself, the level of (13.9) will be relatively high, because $p(E_2/E_1)$ will be relatively low (*i.e.* not much higher than $p(E_2)$), and correspondingly the degree of corroboration will be greater.

A fourth kind of objection to a Bayesian approach runs as follows. The clinician needs statistical data that are relativised to the same relevant demographic and clinical attributes as the patient has. But groups with just this combination of attributes would in many cases be too small for anyone to be able to make accurate estimates of the probabilities in question by reliable procedures. Differences in nosological classification between one time or place and another tend to compound this difficulty, and the difficulty is in any case obviously greater where the disease is a rare one or has bizarre complications or the patient suffers from a chance combination of diseases. And, on the other hand, where adequate statis-

tical data are available for estimating $p(E/D)$, $p(D)$ and $p(E)$, one might reasonably expect them to be available also for estimating $p(D/E)$. So, even where an accurate Bayesian measurement seems possible, it could well be superfluous. This is not a general argument against Bayesianism, since for some A and B, $p(B/A)$ may be much easier to estimate than $p(A/B)$–for example, where A is a scientific generalisation and B reports the successful result of a particular experimental test on it. But, if we know enough to estimate the relative frequency of certain symptoms, signs, *etc.* among those who undoubtedly have a certain disease (presumably because we have eventually identified the latter by chemical tests, autopsies, *etc.*), we presumably also know enough to estimate the relative frequency of those who undoubtedly have a certain disease among those who exhibit the appropriate symptoms, signs, *etc.*

Perhaps a Bayesian may be tempted to rejoin that a reasonably unbiased sample of people with a certain disease may be expected to become known to members of the medical profession, since even those without appropriate symptoms or signs will eventually be revealed by autopsies, whereas an unbiased sample of people with certain symptoms or signs is more difficult to find since those who have these but no corresponding disease tend to disappear from medical view. But this would be a poor rejoinder because the strict Bayesian needs in any case to determine the value of $p(E)$. Also there is an inherent risk of bias in the fact that $p(E/D)$ would normally be determined from statistics of patients referred to hospitals while $p(D/E)$ needs to be evaluated for the population at large including those whose illnesses are sufficiently mild not to require hospitalisation.

A more characteristic rejoinder would be to deny altogether that a Bayesian account of medical diagnosis assumes the availability of adequate statistical data. If a personalist, or subjectivist, conception of probability is coupled with a Bayesian strategy for calculating with probabilities, the difficulties about statistical data are completely sidestepped. The physician who evaluates $p(D/E)$ is now supposed to be doing so in virtue of the several betting-odds he would be inclined to accept, on the basis of all his experience prior to examining the patient, for the patient's having the disease D, for the patient's exhibiting the symptoms, signs *etc.* E, and for the patient's exhibiting E given D, where the physician's overall distribution of bets is sufficiently coherent–*cf.* Kyburg Jr. and Smokler, 1964, p. 11. For example, to accept betting odds of 4 to 1 on the patient's having the disease D is supposed to be tantamount to accepting a value of $\frac{1}{5}$ for p(D).

But what is meant by saying that the physician's previous experience is the 'basis' for his selection of betting-odds here? This selection can-

not be said to be *justified* by previous experience without reintroducing all the above-mentioned difficulties about inadequate statistical data. So apparently the physician's selection of betting-odds is 'based on' his previous experience only in the sense that the former is the causal product of the latter. And patients may well be a little concerned if their physicians can put up no rational defence when the probability of a diagnostic conclusion is challenged. Certainly it will not do to say here, as is sometimes said by personalists in other fields, that in the long run, with adequate exchange of information, the betting-odds of all rational men must coincide. For a patient may well need diagnosis and treatment within a very short run. Nevertheless, a physician's opinions are in practice often respected just because of his reputation for judicious diagnosis; and a personalist Bayesian might claim that this illustrates how we have to fall back on private, inarticulate, 'black-box' processing by a particular good brain wherever rationally defensible estimates cannot be founded on publicly available statistics. Thus the appeal to an expert's clinical judgment, which is so familiar in actual practice, can be given a theoretical representation within the personalist version of Bayesianism. But the validity of this personalist representation will be restricted to those contexts where the actual existence of the disease can in principle be determined in some other way than by reference to the symptoms, signs, *etc.* that constitute the evidence for the diagnosis. A rational bettor does not make bets that cannot be settled. So a personalist Bayesian has to assume that some eventual revelation of the truth-whether by autopsy or by surgery or by paraclinical tests-could in principle take place at some time after the observations on which he has implicitly based his own bet.

A fourth objection to the Bayesian account is more powerful, however, because it affects any method of evaluating the strength of a medical diagnosis in terms of mathematical probability. The objection concerns the importance attachable, at least from the patient's point of view, to the weight of evidence. For example, a physician may attach 95 per cent probability to the patient's having D on the evidence that he in fact manifests E_1. Then, if the physician's judgments are good ones, and a 95 per cent success-rate in diagnosis is all that his conscience, his patients, his profession or his employers require from him, he seems to be justified in diagnosing D from E_1. But more tests may reveal more evidence, E_2, and perhaps $p(D/E_1\&E_2)$ is a great deal less than 95 per cent. So somehow or other the evaluation of a diagnosis has to be related primarily to the extent of the relevant evidence that has actually been obtained. A thoughtful clinician recognises that his reasoning may sometimes be incorrect and that his rate of error may approximate, say,

5 per cent, but he doesn't deliberately bound the thoroughness of his diagnostic examinations by the complacent acceptance of a particular rate of error. A scheme of diagnostic strategy has to be geared to the management of the unusual or atypical situation as well as the usual one, to the 5 per cent of cases that are abnormal as well as to the 95 per cent that are normal. A diagnostician at least *aims* at discovering the truth in each and every case, even if he knows that he can hardly expect actually to discover it in each and every case.

The force of this objection may be brought out by an analogy with legal justice. The criminal courts too can hardly expect to pass correct verdicts in every case. Indeed they would do very well if 95 per cent of their verdicts were correct: there is always the risk of missing or perjured witnesses, incompetent advocates, perverse or corrupt juries, *etc.* So suppose the only evidence presented to the court in the trial of a man accused of possessing illegal drugs is that he possesses a rainbow-painted car and that 95 per cent of those who possess rainbow-painted cars also possess illegal drugs. Would it be justice to condemn him? Should he be deprived of his liberty? No. But why not? Because there isn't *enough* evidence. Further enquiry might show that there was absolutely no trace of illegal drugs anywhere in his house or car. Correspondingly we can't suppose a criminal court to incorporate its diagnostic ambition-the ambition to be right in 95 per cent of its cases-within the standard of proof that it adopts. For, if all that was required was a 95 per cent probability, the man with the rainbow-painted car would be justly condemned. Instead the courts must aim at a correct verdict in every case, and that means that they need to be presented with as much as possible of the relevant facts. If the accused is to be condemned, his guilt must be proved beyond reasonable doubt. So the trier of fact in a criminal court is in a situation that is analogous in some respects to that of a diagnostic physician (even though medical diagnosis does not presume that a patient is innocent of a particular disease unless his having it is proved beyond reasonable doubt). The trier of fact, like the physician, cannot base his verdict *merely* on a high mathematical probability. No doubt the measurement of mathematical probability is sensitive to the impact of corroborative or converging evidence, as we have already seen. But without looking to the extent–or 'weight', as Keynes, 1921, p. 71 ff. called it–of the evidence that we do at present have we cannot tell how much significance to attach to the fact that it happens to converge towards a very high mathematical probability for such-or-such a conclusion.

3.

Let us now look at the main theory of medical diagnosis that is offered as a rival to Bayesian modes of evaluation: cf. especially Kleinmuntz, 1968 and Feinstein, 1974. We shall see that this theory has a complementary virtue, but also a complementary vice. It has the virtue of making the ordinal or comparative grading of evidential weight quite easy. But it assumes, rather optimistically, that in the end the evidence would normally be complete and wholly certain diagnoses would normally become available, even though considerations of time or money may in practice often force the process of collecting evidence about a particular patient to stop before it is complete.

According to this theory diagnostic reasoning may be represented either by a decision-table or by an equivalent flow-chart. In the decision-table representation a series of questions are tabulated, each being susceptible of a yes-or-no answer. An appropriate diagnostic conclusion is then assigned to each familiarly encountered combination of 'yes' and 'no' answers. A flow-chart representation carries the same total amount of information, but may economise in the amount of symbolism that is needed to express it. Its non-terminal boxes are labelled with questions, and two arrows-one marked 'yes' and the other 'no'–lead off from each such box to other boxes: the terminal boxes are labelled with appropriate diagnostic conclusions. So a flow-chart representation can economise with symbolism if several different combinations of answers justify the same conclusion or if one particular answer to one of the questions recurs in the evidence for a number of different diagnoses. A decision-table gives us a conspectus of the detailed variety of obtainable information, with corresponding conclusions, while a flow-chart represents best the variety of different possible sequences of reasoning. A single vast decision-table, or flow-chart, could represent the total diagnostic potential of current medical science, with every stage of reasoning made fully explicit. But in practice one might suppose that a number of much shorter and more specialised diagrams would have greater utility, such as the flow-chart for examining part of a barium enema that was given in News, 1966, p. 36, or the hundred or more flow-charts that have been published in *Patient Care* in recent years.

In practice, because of the risks involved in delay, therapy of some kind may have to be commenced long before all the required information is available and thus before correct therapy is uniquely determined. Or perhaps it would be prohibitively expensive to carry out all the laboratory tests that are required to provide ideally complete information in a particular case. Unless there are special risks in doing so, the obvious

course then is to select the diagnosis that seems most probable on the available evidence. Equally obviously, the merit of this selection depends not only on the relevant degree of mathematical probability, but also on the actual extent to which the questions tabulated on the decision-table have been answered, and thus on the extent to which the relevant route on the flow-chart has been followed through. Measures of probability are familiar; but how do we evaluate the merit of a diagnostic conclusion in terms of the extent of available information?

Writers on medical diagnosis are normally more interested in the heuristic aspect of the reasoning involved here than in its probative aspect. That is to say, they are more interested in solving the methodological problem 'What is the rational way to discover a patient's disease?' than in solving the logical problem 'What is the rational way to justify a particular diagnosis?' But the problems are intrinsically interconnected. A diagnosis reached in the methodologically correct way is justified thereby, and an evaluation of the extent to which a particular diagnosis is justified is ipso facto an evaluation of the extent to which a patient's disease has indeed been discovered. So it is of methodological as well as logical interest to determine the exact nature of the mode of evaluation that is appropriate to decision-table or flow-chart representations of diagnostic reasoning.

My contention here will be that this mode of evaluation is essentially the same as that which Cohen, 1977, p. 121 ff. calls inductive, or Baconian, probability. The point is that the reliability of a generalisation, in experimental science as also elsewhere, is best graded, as Bacon, 1620 saw, by the variety of potentially relevant circumstances that fail to falsify it. (Relevance in this context is established by the successful falsification of generalisations about kindred topics.) The clearest examples of such Baconian gradation are to be found where the reliability of a generalisation is gauged by the results of a controlled experiment and is taken to be higher or lower in accordance with the number of relevant conditions that are controlled in the (successful) experiment. For instance, the curative value of a drug might be established by experiments in which patient's medical histories, ages, diets, *etc.*, were all controlled. The more comprehensive the test that a generalisation survives, the greater the reliability that is ascribable to it. Bacon rightly thought of such tests as putting general questions to nature, so that as the interrogation increases in thoroughness its outcome increases in certainty. It follows that, if a generalisation would be falsified under conditions in which the circumstance V_1 were present, it may be protected against such falsification and thus achieve a higher grade of reliability if it is qualified accordingly. That is, in appropriate cases 'Anything that

is A and V_1 is also B' will be more reliable than 'Anything that is A is also B', because the presence of V_1', excludes that of V_1. (*E.g.* we might ascribe curative value to the drug only when the patients are young.) Correspondingly an inference to a thing's being B from the premise that it is both A and V_1 is stronger (or weightier) than an inference to that conclusion from a premise stating only that it is A. Similarly an inference from the premise that it is both A and V_1' and V_2' may be stronger still. Consider then those pathological generalisations that assert any person to have such-or-such a disease if he has such-or-such symptoms, signs, biographical circumstances and/or paraclinical test-results. Clearly at least one way of grading the reliability of such generalisations is the Baconian method that has just been described. And correspondingly, if the basic research has been done satisfactorily and the textbook generalisations contain all the right qualifications, a diagnosis that is made in conformity with these generalisations will vary in strength (or weight) according as the evidential premisses of the diagnosis for a particular patient include mofe and more of the circumstances that are mentioned in the antecedent clause of the appropriate generalisation.

Ideally we should be able thus to rank-order any given set of diagnoses within a particular field of medicine. In order to guarantee comparability of strength in every case, however, between one diagnosis and another within a particular field of medicine, we need to assume a uniform sequential ordering for the process of collecting evidential data. Such an ordering is implicit in the decision-table or flow-chart representation of diagnostic reasoning. It is also a standard requirement for the list of relevant variables that is to be manipulated in controlled experiments if comparative grades of Baconian reliability are to be established by these experiments. Obviously, there is much to be gained by varying just the most important factors–*i.e.* the circumstances with greatest falsificatory potential–if shortage of time or resources allows only some factors to be varied. The quicker we can get false hypotheses out of the way, the better. So in an idealised reconstruction of diagnostic reasoning, as of experimental research, one would naturally expect an ordering in which the earlier stages eliminate more hypotheses than the later ones. Disease manifestations, as Jacquez, 1964, p. 31 f. put it, need to be given a hierarchy of values. As a result the different kinds of data that are in fact collected for diagnostic purposes in a particular area of medicine standardly constitute a Guttman scale, in the sense that, if we know how many different kinds of data are present in the evidence actually obtained for a particular diagnosis, we should be able to name which kinds they are.

No doubt practical difficulties may often intervene to prevent adoption of the theoretically optimal ordering. But we must then be prepared to accept consequential restrictions on the comparability of results. These constraints are not an argument against accepting a Baconian *logic* for judgments of diagnostic weight. They merely affect the extent to which a flow-chart can give a uniformly applicable account of diagnostic *methodology.*

Now, if it is correct thus to interpret the strength or weight of a diagnosis as a Baconian probability, when the diagnosis is carried out in accordance with the decision-table or flow-chart model, we can use what is known about the logic of Baconian probability in order to derive corresponding properties for this concept of diagnostic weight. In that way we can clarify what constraints one such evaluation of weight imposes on another, just as the mathematical calculus of chance articulates what constraints one Bayesian evaluation of probability imposes on another. And since a generalisation of the Lewis-Barcan modal system S4 articulates the logical syntax of statements about Baconian probability, as in Cohen, 1977, p. 229 ff., we shall find there also a logic of statements about diagnostic weight.

For example, a conjunction of diagnoses acquires as much weight from given evidence as either diagnosis has separately, if they have equal weight, or as the less weighty one has, if they are unequal. So the diagnostic output of a physician does not lose its overall strength as he sees more and more patients. And obviously this is a feature that distinguishes the weight of diagnostic statements very sharply from their mathematical probability, since the multiplicative principle for the probability of a conjunction must entail some reduction in the probability of the overall correctness of a physician's diagnostic output, as he sees more and more patients. It will be less and less *probable* that he got the diagnosis right in every case, even when each patient has precisely the same symptoms, signs *etc.* Yet there is also a sense in which, if every constituent diagnosis is as solid as the evidence permits, then the diagnostician's total achievement is also just as solid. This sense is captured by the Baconian analysis of diagnostic weight.

Again, if the evidence is reliable and gives weight to a particular diagnosis, though the diagnosis is not conclusively proved, a Baconian analysis requires us to say that no weight at all is given to the denial of the diagnosis. And here again there is an important disanalogy with the logic of probability judgments, since if the probability of a diagnosis falls short of complete certainty, the probability of its negation must be greater than zero. The point is that if the balance of evidence builds up in favour of diagnosing that the patient has disease D, the weight of

this build-up may be greater or less, depending on how much relevant evidence has been collected, but since the balance of evidence favours the presence of D it does not favour the absence of D and a zero-grade weighting has therefore to be given to any denial that D is present. So far as evaluations of diagnostic weight embody judgments of Baconian probability, their logic cannot be mapped in any seriously viable way on to the mathematical calculus of chance. It is not just that the two concepts obey different principles for conjunction, negation, *etc.* The logical gulf between the two concepts is much wider because no function can be constructed out of the resources of the calculus of chance that will properly evaluate the strength or weight of a diagnosis that has been carried out in accordance with the flow-chart or decision-table theory.

This would be true even if strength or weight here were an additive property, as is shown in Cohen, 1977, pp.188 ff. and 224 f. But in fact we cannot treat it as an additive property, because the reliability-grade of the pathological hypothesis from which such a diagnosis derives depends on the power of the hypothesis to resist falsification by cumulative combinations of relevant factors, and this power is not an additive property either. Many pairs of casual factors, for example, have an explosive potential when combined with one another, though they are only of slight significance when occurring apart. Cigarette-glow and petrol-vapour constitute a familiar example of this in everday life, or alcohol and barbiturates in medicine. So a Baconian weight-function cannot quantify the strength of a diagnosis, in the way that a successful Bayesian investigation would quantify its probability. Diagnostic weight submits at best to rank-ordering. The resistance that clinicians sometimes feel to quantifying the strength of their diagnostic convictions has a solid rationale. It may be that some concepts of weight admit a Bayesian analysis. But diagnostic weight is not one of these.

4.

What emerges, therefore, is that a Bayesian and a flow-chart (or decision-table) evaluation ought not to conflict with one another, because when properly constructed and appropriately applied they are concerned with different issues and complement one another. The Bayesian evaluation, or indeed any evaluation in terms of mathematical probabilities, declares how far the evidence favours, and how far it opposes, the suggested diagnosis, *on the assumption that this is all the available evidence.* The flow-chart evaluation, or indeed any evaluation in terms of evidental weight or Baconian probability, reveals just *whether* the balance of the evidence obtained favours the suggested mathematical probability,

or not, and if so *how much* of the potentially available relevant evidence is constituted by the evidence actually obtained. In practice the desirability of having any kind of quantitative evaluation in terms of mathematical probability, as a final outcome of the diagnosis, tends to decline as the Baconian probability, or weight, increases. Where there is no time, perhaps, or no inclination to carry out a thorough examination, the patient may have to be treated on the basis of a positively relevant mathematical probability for anyone with his immediately obvious signs or symptoms. At the other extreme, where a thorough and exhaustive collection of relevant evidential data can be made, including the results of appropriate paraclinical tests, the patient's condition may be so well understood that an evaluation in terms of mathematical probability would be superfluous, because so much of the causal process at work has been established that the probability would be only a negligible distance short of 1 (if indeed any worthwhile estimate of the probability were obtainable for so specific a combination of circumstances). But in between these two extremes there is plenty of room for combining both modes of evaluation. To the extent that further evidential data, which might produce certainty, are still lacking, it may well be necessary to add a judgment of mathematical probability to that of evidential weight. But the significance of such a judgment cannot be properly appreciated without due consideration's being given to the weight of the evidence on which it is based. And this applies not only to straightforward judgments of mathematical probability but also to judgments that evaluate some function of the relevant probabilities like Mycin's 'certainty factor'–*cf.* Shortliffe, 1976, p. 171. In general it is in the patient's interest to require as weighty a diagnosis as the risks of obstructing therapy or provoking complications allow, and to have his physician rely as little as possible on the calculation of mathematical probabilities. Economies of finance and professional effort, however, are obviously best served by the opposite policy.

It follows that computerisation may promote diagnostic efficiency in two rather different ways. So far as Bayesian evaluations are concerned, it is clear that a programme for carrying out the requisite calculations on a computer can save time and effort, where the requisite statistical data are available: *cf. e.g.* Overall and Williams, 1963. So far as Baconian evaluations are concerned, however, it is also clear that programmes can be constructed for carrying out that part of a routine diagnostic enquiry which consists in the collection of information about a patient's symptoms, life-style, environment, *etc.* or in monitoring the patterns revealed by paraclinical tests. Such a programme can ask the relevant questions and absorb the relevant answers in a reliably impersonal manner, with

appropriate feedback to the questioning process–*cf.* *e.g.* Kleinmuntz, 1972 and Gorry, 1976–though there are many problems here that are yet to be overcome, as Schaffner, (forthcoming) makes clear. These two kinds of computerisation are just as little rivals to one another (except perhaps for the use of computer time) as are the two different types of diagnostic evaluation. Finally, one needs to beware of drawing the wrong conclusions from the fact that in a particular series of trials a Bayesian programme may have been as successful in making the right diagnosis as a group of wellqualified clinicians: *cf.* Overall and Williams, 1963. This is exactly what one should expect if the computer is really supplied with just the same *weight* of evidence as the clinicians consciously or unconsciously obtain. It is quite another matter to judge in a particular case, whether that weight of evidence is heavy or light, adequate or inadequate; and for this we need Baconian, not Bayesian reasoning.[1]

Notes

1. I am grateful for some helpful comments on an earlier version of this paper to Jonathan E. Adler, Godfrey Fowler, Kenneth F. Schaffner, and R. G. A. Williams, and to participants at a meeting of the Society for Exact Philosophy in Montreal, Quebec, on June 5, 1979, and at a session of the 6th International Congress for Logic, Methodology and Philosophy of Science at Hannover, West Germany, on Thursday, August 23, 1979.

References

Bacon, F. (1620). *Novum Organum.* London.

Cohen, L. J. (1977). *The Probable and the Provable.* Clarendon Press.

Feinstein, A. (1973 and 1974). An analysis of dagnostic reasoning. *Yale Journal of Biology and Medicine*, 46 and 47:pp. 212–32, 264, 283 and pp.5–32.

Feinstein, A. (1977). Clinical biostatics xxxix. the haze of bayes, the areial palaces of decision analysis and the computerized ouija board. *Clinical Pharmacology and Therapeutics*, 21:pp. 482–96.

Gorry, G. (1976). Knowledge-based systems for clinical problem solving. In de Domball, F. and Grémy, F., editors, *Decision Making and Medical Care: Can Information Science Help?*, pages 23–31.

Jacquez, J. (1964). The diangostic process: problems and perspectives. In Jacquez, J., editor, *The Diagnostic Process*, pages 23–37.

Jacquez, J. and Norusis, H. (1976). The importance of symptom non-independence in diagnosis. In de Domball, F. and Grémy, F., editors, *Decision Making and Medical Care: Can Information Science Help?*, pages 379–92.

Keynes, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

Kleinmuntz, B. (1968). Diagnostic problem-solving by computer. *Japanese Psychological Rsearch*, 7:pp. 189–94.

Kleinmuntz, B. (1972). Medical information processing by computer. In Jacquez, J., editor, *Computer Diagnosis and Diagnostic Methods*, pages 45–72. Springfield.

Kyburg Jr., H. and Smokler, H. (1964). Editorial introduction. In H.E. Kyburg, J. and Smokler, H., editors, *Studies in Subjective Probability.* New York.

Lange, H. (1976). Decision-making by doctors and aided by computers. In de Domball, F. and Grémy, F., editors, *Decision Making and Medical Care: Can Information Science Help?*, pages 11–22.

Ledley, R. and Lusted, L. (1959). Reasoning foundation of medical diagnosis. *Science*, 130:pp. 9–21.

Lusted, L. (1968). *Introduction to Medical Decision Making.* Springfield.

News, M. (1966). Medical news section. *Journal of the American Medical Association.*

Overall, J. and Williams, C. (1963). Conditional probability program for diagnosis of thyroid function. *Journal of the American Medical Association*, 183:pp. 307–13.

Sardeman, W. (1964). The history and physical examination. In Jacquez, J., editor, *The Diagnostic Process.* University of Michigan Press.

Schaffner, K. Problems in computer diagnosis. In Engeldhardt, H. and Spicker, S., editors, *Ethics and Clinical Diagnosis*, Philosophy and Medicine.

Shortliffe, E. (1976). *Computer-based Medical Consultations: MYCIN.* Elsevier, New York.

Tversky, A. and Kahneman, D. (1977). Causal thinking in judgement under uncertainty. In Butts, R. and Hintikka, J., editors, *Basic Problems in Methodology and Linguistics*, pages 167–90.

Wulff, H. (1976). *Rational Diagnosis and Treatment.* Blackwell, Oxford.

14

ARE PEOPLE PROGRAMMED TO COMMIT FALLACIES? FURTHER THOUGHTS ABOUT THE INTERPRETATION OF EXPERIMENTAL DATA ON PROBABILITY JUDGMENT.

1. THE NORM EXTRACTION METHOD

During the last decade or so a characteristic method of interpreting experimental data has been followed by most researchers in their investigations into how statistically untutored subjects actually judge probabilities. For the most part the investigator tacitly assumes that the problem-task set to his or her subjects is correctly soluble only in terms of some academically well-regarded conception of probabilities that he or she has in mind. The investigator therefore evaluates the subjects' performance for correctness or incorrectness by a technique of assessment or estimation that is appropriate to this mode of conception. If some subjects' responses seem, when thus evaluated, to be incorrect, their error may be put down by the investigator–in terms of the computational metaphor–either to a standing fault in the programming of ordinary people (in that they are not programmed to apply appropriate valid principles to the task in question or are programmed to apply inappropriate or invalid ones) or to a temporary malfunction in the running of a valid programme.

We may call this method of interpreting experimental data about probability judgments "the Preconceived Norm Method" and contrast it with what is conveniently called "the Norm Extraction Method". The Norm Extraction Method assumes instead that, unless their judgment

is clouded at the time by wishful thinking, forgetfulness, inattentiveness, low intelligence, immaturity, senility, or some other competence-inhibiting factor, all subjects reason correctly about probability: none are programmed to commit fallacies or indulge in illusions. So by setting subjects a particular kind of problem-task the investigator may at best expect to discover from their responses what conception of probability they implicitly adopt for solving that kind of problem and how they assess or estimate the size of the probability.

In short, the Preconceived Norm Method assumes a standard conception of probability, imputes its acceptance to the subjects, and hypothesises either faulty programming or temporary causes of malfunction in order to account for estimates that are erroneous in terms of that conception: the Norm Extraction Method hypothesises about the subjects' conception of probability and their mode of assessing it, on the assumption that unless affected by temporary or adventitious causes of error their judgments are correct.

The issue between these two methods is not of a kind that lends itself to being determined by a single crucial experiment since it concerns a whole policy or paradigm of research. The principles implicitly involved are not just straightforward generalisations that are open to being refuted by the discovery of counter-instances. Instead they are multiply quantified claims which assert that *all* recurrent patterns of error in lay judgments of probability have *some* explanation of such-or-such a kind (in the one case, of either the faulty-programme kind or the cause-of-malfunction kind; in the other case, only of the cause-of-malfunction kind). *All*-and-*some* principles of this nature often play an important methodological role in the development of scientific research, but because of their logical structure they are inherently neither verifiable nor falsifiable (Watkins, 1958). The success of such a principle may be gauged only by the extent to which it helps to make coherent sense of the already available data throughout the relevant field of investigation and also continues to promote adequate explanations there. So it is essential to consider the issue between the Preconceived Norm Method and the Norm Extraction Method in the context of a wide range of different experiments. Thus in Cohen (1979) and Cohen (1981 a) several examples were given of how allegedly prevalent fallacies about sample-size and prior probabilities may be seen not to be fallacious at all when treated in accordance with what has here been called the Norm Extraction Method: the subjects' allegedly erroneous responses in these cases turn out to be capable of being understood in a way that admits their validity. And in the present paper a number of other well-known results will be discussed in order to show how extensively the Norm Extraction Method

fits the facts. It is now rather doubtful whether any firm evidence will ever be found that human beings are programmed to judge probabilities faultily in certain circumstances (e.g. by the so-called heuristic of representativeness).

Certain *a priori* considerations also favour this conclusion.

One such consideration is that the Norm Extraction Method is more economical in the variety of explanatory factors that it envisages. It admits only normatively correct mechanisms on the one side, and adventitious causes of error on the other. The Preconceived Norm Method admits not only both these types of factor, but also certain invalid heuristics-cognitive routines that sometimes generate correct answers and sometimes do not. This is because, where subjects' responses do not conform to the assumed norm but seem too regular and clearly patterned to be put down to adventitious causes, the Preconceived Norm Method tempts an investigator to suppose that these responses are programmed and therefore to postulate a special category of faulty programmes in order to account for them.

A second consideration stems from the fact that the mathematical calculus of chance has long since been recognised to have more than one possible interpretation as a theory of probability. As Hacking (1975, p. 13) points out, Poisson and Cournot saw, long before Carnap, that this calculus could be understood as providing principles for both an objective and an epistemic conception of probability. Nor, in fact, are only two different interpretations admissible, as Nagel (1955) has shown. Rather we have on our hands a new problem: why are at least five strikingly different interpretations possible? For we can interpret the probability-function that is regulated by the classical calculus of chance either as a ratio of *a priori* chances, or as a relative frequency of empirically given outcomes, or as a causal propensity, or as a logical relation, or as a measure of actual or appropriate belief. Mackie (1973, p. 155ff) sought to explain this in terms of a series of historically natural shifts and extensions in the meaning of the term 'probability'. But this explanation is not only unsupported by the chronological facts. It is also inadequate to the demands of the problem (Cohen, 1977, pp. 11-12). Instead we have to think of probability as a generalisation of the concept of provability and as necessarily then taking a variety of semantically specific forms in accordance with familiar distinctions between different kinds of proof-systems (Cohen, 1977, p. 13ff). Moreover, if for that reason any coherent gradation of provability is a probability, we must also allow the possibility of certain non-Pascalian kinds of probability judgments which do not even conform to all the constraints imposed by the classical (Pascalian) calculus of chance. In particular we can have non-additive

probabilities–probabilities that do not conform to a complementational principle for negation–as Bernoulli long ago recognised (Shafer, 1978) and as the Baconian tradition also held (Cohen, 1980). Just as we can usefully measure quantities of apples in at least three different ways (by number, by weight or by volume), so too, if we so desire, we can measure probabilities in several different ways according to the purpose in hand. A may well be more probable than B by one measure, and less probable by another.

Since therefore more than one conception of probability is possible, an experimenter has somehow to establish which is operative in the reasonings of his subjects about the problem-task presented to them. But how can he do this independently of the actual data that emerge from the experiment? If he recognises the risk of ambiguity and instructs the subjects beforehand in which conception of probability they are to use and how they are to use it, he is no longer dealing with naive and untutored subjects. The subjects would have been tutored in a certain conception of probability, and any errors that they then commit may legitimately be treated as a measure of the inadequacy of the experimenter's instructions rather than as a measure of their own native irrationality. So, if the experimenter sets his subjects a real-world task and does not instruct them beforehand about which conception of probability they are to use, he has to extract from their behaviour what conception they actually use. And this has to be done within a total scheme of interpretation that makes due allowance for the possibility that subjects may make genuine errors in estimating the sizes of particular probabilities that satisfy their conception. Certainly it is unsafe for the investigator to rely on having formulated the wording of the task so unambiguously that it is bound to convey to the mind of each untutored subject precisely the same problem as the investigator himself has in mind (cf. Ebbesen and Konecni, 1980, p. 38, and J. S. Carroll, 1980, p. 70).

A third consideration is also worth mentioning, though opinions may differ about how far it takes us. Where deducibility hinges on the meanings of logical particles or quantifiers, ordinary people cannot be supposed to be systematically incorrect in their intuitions about what is deducible from what. The criteria of correct deduction are implicit in their own speech practice–although what has to be taken into account in each case is the total content of the message communicated and not just its linguistic form (Cohen, 1981a, p. 326f.). Of course, people sometimes contradict themselves or make invalid deductions. But the norms that they infringe when they do so are discoverable from their own intuitions. So to understand them correctly one must for the most part be charitable, and assume that they have not spoken illogically. How

far can an analogous point be made about probability? The trouble here is that one needs to distinguish between this or that conception of probability, on the one hand, and appropriate strategies for assessing or estimating it, on the other.[1] For example, when probability is conceived as a relative frequency, a variety of sampling techniques are available for its estimation, and when probability is conceived as a measure of the strength of justified belief, there is a variety of ways in which coherent schemes of betting-odds may he devised in order to grade this strength in a particular context. And in each of these two cases the conception of probability that is operative can, at least in principle, be identified independently of the method of estimation employed.

But it does not have to be so identified. That is to say, there are two ways of individuating the operative conception of probability. One can think of it rather narrowly, e.g. as a relative frequency to be estimated by such-and-such a sampling strategy. Or one can think of it, more broadly, as a relative frequency however estimated. In terms of the former, narrower mode of individuation, a subject's conception of probability has to be inferred (subject to the possibility of adventitious error) from the nature of the subject's own responses. Hence there can be no room to impute erroneous programming. But even if this strongly apriorist argument is unacceptable, and the broader mode of individuating conceptions of probability is therefore preferred, so that the fundamental question at issue becomes an empirical one, the evidence nevertheless tends in the same direction. It is logically possible, for example, that the ordinary layman supposes small samples to be more indicative than large ones in relation to long-run relative frequencies. But it turns out empirically false that he does so (see sec. VII below).

Charitableness, however, has only a rather limited role to play in the interpretation of experimental data about probability judgment. It would be a highly misleading caricature of the Norm Extraction Method to represent it as a policy of always, or nearly always, interpreting subjects' responses in a sense in which they are correct. Such a policy would indeed be methodologically indefensible: we are not entitled to infer *just* from subjects' methods of assessing or estimating probabilities to their use of a conception of probability for which these are reasonable answers. It is worth emphasising, therefore, that two constraints exclude the Norm Extraction Method from permitting this procedure. First, we have to suppose that the conception of probability in terms of which a subject comes to construe his task is cued for him in each case by the wording and content of his instructions (against the background of his own experience and education and any other individual differences that are relevant to his cognitive performance). Hence the Norm Extraction

Method requires investigators to hypothesise regularities in such cueing which block arbitrarily charitable attributions of validity. Relevantly similar subjects must be expected to operate with the same conception of probability in relevantly similar contexts. Secondly, people do make mistakes quite often, and the causes of error must also be supposed to function regularly. So if an error is imputed to a subject in one case and attributed to a particular cause of malfunction there in order to provide a coherent account of the situation, then that factor must be expected to cause similar malfunctions in all relevantly similar cases.

2. THE CONCEPT OF SUBJECTIVE PROBABILITY

Research-workers in psychology, as in other biological sciences, most naturally think of probabilities as long-run relative frequencies. Such probability-functions take sets as arguments and their values are ratios between the cardinalities of those sets. In this way psychologists can see themselves as dealing with inherently measurable quantities and can thus seek to maintain the characteristically quantitative objectivity of modern science. It would hardly he surprising, therefore, if some of them were unselfcritically inclined to suppose that laymen also think most naturally of probabilities in the same terms. But ordinary people, it appears (see sections III-VII below), do not naturally think of probability in such terms unless appropriately cued to do so. They either think of probability within a non-Pascalian framework, or they prefer a Pascalian probability-function that takes propositions as arguments, in accordance with the so-called "subjectivist" or "personalist" account, or one that takes properties as arguments, in accordance with the "propensity" account. And here it is necessary to point out that the term "subjective probability" is ambiguous, and ambiguous in more than one way.

The first ambiguity arises from a difference between psychological and philosophical usage. It is widely assumed by psychologists that "the subjective probability of X ... is veridical to the extent that its value agrees with the corresponding objective probability" (Peterson et al., 1965, p. 526). Now that assumption is certainly correct, though trivial, where "the subjective probability of X" means "a subject's estimate of the objective probability of X". But it is just as certainly incorrect where "the subjective probability of X" means "the strength of belief in X" and "the objective probability of X" means "the relative frequency of X". When people measure the probability of an individual outcome by evaluating the strength of belief that they have in that outcome in accordance with acceptable odds on its occurrence, as proposed by Ramsey (1931), De

Finetti (1931), Savage (1954), and Kyburg and Smokier (1964), they can take into account not only what they know about relevant relative frequencies but also what they know about anything else, including the extent of their knowledge about relevant relative frequencies. The probability that Jones will survive to the age of 70, on the evidence that he is a university teacher, depends then not just on the relative frequency of such survivals among university teachers but also on the extent to which other relevant information about Jones (his medical history, leisure activities, etc.) has not been included in the evidence. The odds on Jones's dying before the age of 70 might be quite long if the only relevant fact that needed to be examined were his mode of earning a living and he was known to be a university teacher. But if only that fact were known, whereas there were several other relevant questions about Jones that remained uninvestigated, a bookmaker would shorten his odds in order to reduce his risk. The possibility that, in this sense, subjective probabilities–or "personal probabilities", as it will be less confusing to call them–may differ legitimately in value from the corresponding objective relative frequencies turns out to be vitally important for the interpretation of experimental data on probability judgment.[2] Indeed, in view of the fact that many Bayesian theorists (Savage, 1964) have combined their preference for Bayesian modes of calculation with a preference for the personalist conception of what they are calculating (or have even thought that the latter preference provided a rationale for the former), it is quite surprising to find that many psychologists are inclined to combine respect for Bayesian norms with a tendency to ignore the legitimacy of the personalist conception. Only thus has it seemed at all plausible for them to decry "the fallacy of conservatism in information integration", "the gambler's fallacy", "the fallacy of evaluating probabilities by sample-to-population ratio", and so on. Of course, it is possible that these psychologists all reject the possibility of measuring probabilities in the way suggested by the Ramsey-De Finetti-Savage school. But, though the measure in terms of coherent betting-odds may be inapplicable in certain kinds of cases, such as issues about the past where no more evidence can be obtained (Cohen, 1977, pp. 26-27, and 1981b), its mathematical possibility is scarcely in doubt. Or at any rate the rejection of that possibility needs to be supported by new argument, and no such support is to be found in the relevant psychological literature.

But there is also another important ambiguity in the notion of subjective or personal probability. When a person judges the probability of X by the strength of his belief in X, does his statement of X's probability mean just that he himself, on the evidence, believes in X with that

strength of conviction? Or does it mean that anyone, on the evidence, ought to believe in X with that strength of conviction? The former (autobiographical) version of the personalist account has a feature that distinguishes it from all other analyses of probability judgment: two people who ascribe different probabilities to the same outcome are not contradicting one another, since each is merely describing his own mental state. But the other (deontic) version of the personalist account lacks this idiosyncratic feature and is correspondingly more widely applicable. Indeed, if it can be construed as measuring by appropriate ratios wherever betting odds are inappropriate, it could be thought of as differing from the analysis of probability as a causal propensity only when looked at from a rather metaphysical point of view.[3] Wherever both analyses operate, both should assign the same probability-values to the same outcomes, since the propensity analysis allows us to take into account the extent to which the influence of Jones's profession on his survival might be affected by other unstated factors just as the personalist analysis allows us to take into account the extent to which other information about Jones might be relevant. But the propensity analysis will be more congenial to those who prefer to think that their judgments about probability are describing something that is objectively real, while the personalist account will be more congenial to those idealists, empiricists and others who prefer to confine the real to the observable. Fortunately the subjects in the experiments to be discussed need not be supposed to concern themselves with this metaphysical issue.

3. CONSERVATISM IN INFORMATION INTEGRATION

Let us now explore some of the consequences of applying the Norm Extraction Method instead of the Preconceived Norm Method. We shall see that the concept of personal probability, or its objectivist counterpart, has an important part to play.

Consider first the alleged fallacy of conservatism, which has been authoritatively claimed (Slovic, Fischhoff and Lichtenstein, 1977) to be "the primary finding of Bayesian information integration research". In certain easily reproduced circumstances, it is claimed, laymen do not make big enough revisions of their probability estimates when presented with additional information. For example, Phillips and Edwards (1966) had their subjects estimate whether a randomly chosen bag contained predominantly blue or predominantly red chips, by–in effect–drawing a chip at a time from the bag and replacing it. The subjects were told the ratio of bags containing predominantly blue chips to those containing

predominantly red ones, and this was assumed to afford them a prior probability. They were also told the number of chips of the majority and minority colours, which were the same in each bag. A sequence of twenty chips was shown them, drawn one at a time from the chosen bag and each replaced immediately after being shown. And the subjects were asked to revise their previous intuitive estimates (of the probability that each type of bag had been chosen) after each new chip was shown. In the event the subjects revised these estimates less than the amount that the investigators calculated they should have done according to Bayes' law. So the investigators concluded that conservatism in the integration of new information was a pervasive effect, increasing as the diagnostic value of a single chip increased. Similar results have been reported by several other investigators such as Peterson and Du Charme (1967), Messick and Campos (1972) and Donmell and Du Charme (1975), despite the fact that this allegedly excessive respect for prior probabilities seems difficult to reconcile with subjects' alleged tendency, reported by Kahneman and Tversky (1973), Lyon and Slovic (1976) and Bar Hillel (1980), to ignore prior probabilities altogether.

Is there a real fallacy here? Of course, a fallacy has to be admitted if we assume, in accordance with the Preconceived Norm Method, that the subjects are conceiving the probability of an outcome as the ratio of the number of favourable chances to the total number of possible chances. On that conception the relevant likelihoods–e.g. the probability of drawing a red chip (D), on the hypothesis that the bag contains 70 red chips and 30 blue ones (H)–are strictly determined by the terms of the problem. "Simple statistical theory" (Alker and Hermann, 1971, p. 134) suffices to settle these likelihoods. And the correct revisions are then immediately apparent from an application of Bayes' theorem, since the prior probability of H has been stated and the prior probability of D can be calculated from the information given.

But the subjects are surely entitled to be a little cautious about just assuming that the selection of a chip of particular colour from a bag is a matter of pure chance. Who knows how the bag was filled or shaken, or what the effect this will have on drawings from it? "Simple statistical theory" may not be the right guide to follow here.[4] The layman's black box should be expected to be as pedantic as a computer in not assuming input that it has not had. So let us pursue the Norm Extraction Method instead and seek an interpretation of the subjects' probability function that is more in keeping with their state of information and the content of their responses. Suppose that the subjects conceive probability as a measure of appropriate intensity of belief. They will not then treat the hypothesis about the bag's composition as being quite so relevant,

whether positively or negatively, to the probability of picking a chip of a particular colour. $p(D, H)$ will therefore be rather closer to $p(D)$ than it would be on the assumption that pure chance is operating, so $p(D, H)/p(D)$ will be nearer to 1. Therefore by Bayes' theorem, which says that (where $p(D) > 0$)

$$P(H, D) = \frac{p(D, H) \times p(H)}{P(D)}$$

$p(H, D)$ will be closer to $p(H)$. In other words the fallacy of conservatism here is an artefact of the Preconceived Norm Method. If instead we apply the Norm Extraction Method, we can interpret the experimental data in a way that allows us to infer the subjects to have been consistently Bayesian in their calculations.

Admittedly Vlek and Van der Heijden (1967) claim to have found that for individual data, direct posterior probability estimates are not systematically related to values inferred from likelihood estimates. But this finding seems to be contradicted by results that Peterson et al. (1965), Beach (1966), Messick and Campos (1972) and Vlek (1973) describe.

Two other experimental findings help to support the proposed interpretation.

First, if so-called "conservatism" in the face of present evidence is really the result of conceiving probability as the right strength of belief rather than as the ratio of favourable chances, one would expect that any encouragement of subjects to conceive of probabilities in the latter terms would diminish or eliminate conservatism. And so it turns out (Messick and Campos, 1972): calibration in such cases destroys conservatism.

Secondly, if this so-called "conservatism" resulted from some inherent inadequacy in people's information-processing systems one might expect that, when individual differences in information-processing are measured on independently attested scales, some of them would correlate with degrees of "conservatism". In fact no such correlation was found by Alker and Hermann (1971). And that is just what one would expect if "conservatism" is not a defect at all, but a rather deeply rooted virtue of the system.

Other experimental findings in this area seem too uncertain to be taken into account. For example, when data confirm a subject's hypothesis, "conservatism" is sometimes reported as being increased (Du Charme and Peterson, 1968) and sometimes as being decreased (Pitz, Downing and Reinhold, 1967). It is difficult to understand this difference unless it is some kind of experimental artefact. Similarly, subjects are said to get even more conservative as the perceived importance and

complexity of their task increases (Alker and Hermann, 1971). But unfortunately in just these cases it becomes riskier to take for granted that the subjects' judgments of prior probabilities analogous to $p(D)$ or $p(H)$ above are psychologically independent of their judgments of likelihoods analogous to $p(D, H)$: "it is precisely the amount of independence that may be changing as problems become less like mathematical puzzles and more like 'real-life' decisions" (Alker and Hermann, 1971, p. 39). Finally, there is some evidence that "conservatism" may increase in later trials (Du Charme and Peterson, 1968) though this effect is not reported by Phillips and Edwards (1966). If the effect is genuine, it would just suggest an increasing reluctance on the part of subjects to assume that the data are being obtained by a random or chance proces.

4. THE GAMBLER'S FALLACY

The same manoeuvres–replacing a ratio-of-chances interpretation of the subjects' probability judgments by a personalist or propensity one–will serve to transform the so-called "gambler's fallacy" into a valid conclusion from a trio of premisses, each of which is inherently reasonable in the situation concerned. Investigators who ignore this (e.g. Tversky and Kahneman, 1971) do so at the risk of misinterpreting subjects' responses. Of course, where the probability of any single outcome in a game of chance is measured by the inverse of the number of severally exclusive and jointly exhaustive outcomes, it is a fallacy to suppose that the probability of one particular outcome on the present occasion is in any way affected by the actual outcome on any other occasion. If probability is measured in terms of *a priori* given alternatives, as is appropriate for undoubted games of chance, then ex hypothesi no physical event or empirical fact can possibly be relevant to the determination of the probability. For example, if we think of coin-tosses in these terms, then, however long a run of heads has occurred with a particular coin, it is a fallacy to suppose that the probability of the coin's falling tails on the next toss is anything less, or greater than, $\frac{1}{2}$. But the situation is quite different for a person who does not take as certain that the game is one of pure chance and therefore measures the probability instead by appropriate intensity of belief or strength of natural propensity. For such a person it becomes reasonable to take into account any relevant empirical facts, such as the balance of the coin and its toss, or the long run of heads, and each of the following three premises becomes inherently plausible. The derived conclusion then follows by a series of deductively valid steps. If the three premises are true, the conclusion must also be true:

Premiss (i). It is highly probable (i.e. much more probable than not) that every toss's outcome is a matter of chance. (i.e., the presumption is that the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome.)

Premiss (ii). if it is highly probable that every toss's outcom's outcome, that outcome will probably not increase the present low probability that the coin, or the method of tossing it, is biased. (I.e., if you're strongly inclined to believe that the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome, then you're inclined to believe of each future toss's outcome that it will not reduce your inclination to believe the causal factors controlling the outcome of a toss cannot be determined sufficiently to found a prediction of the outcome; or, more generally, if you're strongly inclined to believe that p, then you're inclined to believe of each as yet unknown piece of evidence that it will not be unfavourable to p.)

Premiss (iii). if the next outcome, after the present long run of heads, is heads yet again, that outcome will increase the present low probability that the coin, or the method of tossing it, is biased. (I.e., despite the presumption in (i), only a born sucker–or a victim of the sorites paradox–never suspects a rigged game.)

(iv). in the case of each toss's outcome that outcome will probably not increase the present low probability that the coin or the method of tossing of it is biased (by modus ponens from (i) and (ii)).

(v). if the next outcome will not increase the present low probability that the coin, or the method of tossing it, is biased, then that outcome, after the present long run of heads, will not be heads yet again (by contraposition from (iii)).

(vi). if it is probable that the next outcome will not increase the present low probability that the coin, or the method of tossing it, is biased, then that outcome, after the present long run of heads, will probably not be heads yet again (from (v) in virtue of the rule "From X's being a consequence of Y infer $p(X) \geq p(Y)$").

Conclusion (vii). it is probable that the next outcome, after the present long run of heads, will not be heads yet again (by universal instantiation and modus ponens from (iv) and (vi)).

This argument may encounter the objection that any such subject who offered short odds on tails in accordance with the conclusion (vii) would be bound to lose money in the long run. But would he? The only apparent irrationality in such a bettor's conduct is that while offering short odds on tails he also seems committed to offering short odds on the truth of the proposition that evens are the correct odds on tails, since by (i) he assumes it probable that each toss's outcome is a matter of chance. And to live up to this latter commitment would in fact be a prudent and coherent way of laying off any bets made in accordance with (vii). A person who loses money in the long run on his bets in accordance with (vii) could win back his losses by betting, in accordance with (i), that he was accepting wrong odds in accordance with (vii). So the only way to challenge the argument seriously would be to prove that one or more of the premisses is fallacious. But that does not look like being an easy task. Certainly one should not think ill of a layman who asserts (vii), since his assertion can be rationalised as a deduction from inherently plausible premisses (even if he himself is unaware of this).

There may also be other ways of rationalising the gambler's fallacy (Cohen, 1981a, pp. 327-328), and no doubt many people get sufficiently muddled about probability to make genuine mistakes about it. In order to find out exactly what goes on, and when, we need more experimental data, with subjects' protocols. But there is at present no simpler potential explanation of the prevalence of the so-called gambler's fallacy than that for the most part it is not a fallacy at all and that all we can tell by a gambler's attitude of approval or disapproval towards the "fallacy" is whether he is measuring his probabilities by strength of belief (or natural propensity) or by a priori ratios, respectively. Not surprisingly, indeed, it is found that the negative recency effect exhibited in the gambler's fallacy tends to disappear after a large number of trials (Lindman and Edwards, 1961). That is to say, as calibration becomes possible premise (iii) above becomes less plausible.

Certainly the Preconceived Norm Method has not come up so far with any acceptable explanation. Those who assume that the probabilities concerned are *a priori* ratios, and that therefore the gambler's "fallacy" is a real error, have to explain why people are so prone to commit this error. They are therefore led to attribute the error to the alleged prevalence of some fallacious heuristic, such as the so-called heuristic of representativeness. For example, Tversky and Kahneman (1971, pp. 105-106), followed by Nisbett and Ross (1980, p. 25), tells us that the gambler's fallacy is to be accounted for by such facts as that the occurrence of black on a roulette wheel after a long run of red is more representative of the generating process than the occurrence of another

red. Yet the occurrence of black after only one occurrence of red, or of two blacks after two reds, would be even more "representative". So this theory leaves quite unexplained why it is that real gamblers are reputed to expect black after a long run of red but not after a short run, even though the latter would be more "representative" than the former. Nor are they reputed to expect a long run of black after a long run of red, though again this would be more representative. On the other hand, as we have seen above, if we follow the Norm Extraction Method these facts can be quite satisfactorily explained without our invoking the prevalence of any potentially fallacious procedure like the representativeness heuristic.[5]

5. JUDGING THE PROBABILITY OF A CONJUNCTION

Another profitable field for applying the Norm Extraction Method is in regard to experiments about judging the probability of a conjunction.

In a recent study Kahneman and Tversky (forthcoming, cf. Nisbett and Ross, 1980, p. 146) gave subjects personality profiles of various target persons. Subjects were then asked to assess the likelihood that the persons described in the profiles belonged to various groups. One group of subjects was asked to estimate the likelihood that persons with given profiles were members of non-compound groups like lawyers or Republicans. Another group of subjects was asked to estimate the probability that the profiled persons were members of compound groups like Republican lawyers. What Kahneman and Tversky found is that if a profiled person is judged rather unlikely to be, say, a lawyer, and rather likely to be a Republican, he will be judged moderately likely to be a Republican lawyer. So the investigators concluded that their subjects had committed the fallacy of supposing that the mathematical probability of a person's being a Republican lawyer is higher than that of his being a lawyer and they explained this by attributing to their subjects a heuristic of representativeness that was programmed in such a way as to commit this fallacy in such cases.

But, if we follow the Norm Extraction Method and assume instead that most subjects were not committing any fallacy here, it is easy enough to infer an appropriate interpretation of the data that pays due regard to how the subjects might be expected to view their task. All that is needed is to infer that, in keeping with familiar findings of attribution theory, the subjects have taken the question to concern the causes or effects of the collections of features set out in the profiles and are out to assess the believability of a story about an individual that is cast in

these terms, rather than to estimate the relative frequencies of different kinds of people. For it would be quite coherent to think of these features as rarely resulting just from a person's being a lawyer, very often from a person's being a Republican, and quite often from a person's being a Republican lawyer (because the effect of being a lawyer is to weaken, but not to destroy, the relevant causal efficacy of Republican Party membership). Or the features might quite coherently tend to encourage a person to become a Republican and to deter him from becoming a lawyer, unless he in fact becomes a Republican. Of these two possible attitudes towards the profiled features, the former seems inherently more plausible, since it treats the lawyer *versus* Republican question as a request for a diagnosis, as it were, of the symptoms present in the profile. But both attitudes can easily be reconstructed either within a personalist or propensity system of Pascalian probability or within a Baconian system of noncomplementational probability that is appropriate for reasoning about causal connections (Cohen, 1979). So there is no need at all to suppose an intrinsically fallacious mechanism here.

Again, we are told (Nisbett and Ross, 1980, p. 147) that subjects estimate the probability of a compound sequence of events of a certain kind, on given evidence, to be greater than the least likely of the events in the sequence. But the interpretation appropriate for this datum is rather like that appropriate for alleged commissions of the fallacy-of-the-converse (Cohen, 1981a, pp. 326-327). One has to take into account not just the meanings of the sentences in which instructions to subjects are formulated but also the implications of uttering them. So when asked for the probability of a particular single event subjects may well infer that what is wanted is an estimate of the believability of that event's occurrence as an apparently isolated effect, which could well be lower than the believability of the occurrence of a particular causal sequence containing the event. If, on the other hand, subjects are asked specifically for the probability of the single event's occurrence *whether* in isolation *or* within the particular sequence, it would be very surprising indeed if they then went on to declare the particular sequence's occurrence to be even more probable.

The Norm Extraction Method also has an important bearing on that type of experiment with drawing marbles from urns in which significant numbers of subjects are alleged to overestimate the probabilities of conjunctive outcomes and underestimate the probabilities of disjunctive ones (e.g. Cohen and Hansel, 1957; Bar-Hillel, 1973). This interpretation of the data results from applying the Preconceived Norm Method, because it is assumed that all the subjects share the investigator's conception of the act of drawing a marble out of a mixture of white

and coloured marbles in an urn, with replacement, as a perfect game of chance. If that assumption were correct, then each act of drawing a marble (from an urn containing stated proportions of white and coloured marbles) would be thought independent of every other such act. And no doubt some of the subjects, who give what the investigators think to be the correct answers, do indeed conceive the act of drawing a marble in just this way. But the others, who give the allegedly incorrect answers, may according to the Norm Extraction Method be supposed to measure their probabilities in a way that allows their estimates to be correct. Specifically they may be supposed to measure their probabilities by the intensity of their beliefs and therefore, because of the measures they assign to compound events, to regard the draws as not wholly independent of one another. Can we be quite sure, they might be supposed to ask themselves (if they thought about it explicitly), that when a coloured marble is drawn first it may not be evidence of a tendency in the mixing process to bring coloured marbles within easier reach, or that when a coloured marble is replaced it may not tend to increase the number of coloured marbles that lie most readily to hand within the urn? Is not therefore the probability of drawing two coloured marbles slightly higher than the square of the probability of drawing a single one? Indeed, when one reflects on the matter, it is really quite curious that investigators, both here and in the "conservatism" and "gambler's fallacy" cases, should have preferred to impute a mathematical fallacy to many of their subjects rather than just to impute an implicit element of doubt whether the experimental apparatus constituted a perfect game of chance. After all, even a very long series of tests could establish no better than a high probability that it was actually a game of chance.

Perhaps subjects' protocols could usefully be sought in such experiments in order to discover whether they confirm the presence of doubt about its being a game of pure chance. It would certainly be interesting to have such protocols. But no protocols could confirm the view that laymen are inherently prone to overestimate the probabilities of conjunctions or underestimate those of disjunctions, since it is quite clear that when the task is unambiguously presented lay subjects are capable of responding in accordance with correct mathematical principles in regard to the Pascalian probability both of a conjunction (Beach, 1966) and of a disjunction (Beach and Peterson, 1966).

6. THE SIGNIFICANCE OF SAMPLE-TO-POPULATION RATIO

Another issue that deserves examination here concerns proportionality in sample size. It seems well established (Bar-Hillel, 1979) that under certain circumstances subjects attach evidential significance to the ratio of the size of a sample to the size of its population, irrespective of the absolute size of the sample. For example, if 1000 voters have been polled in each of two cities, one with a population of 50,000 voters and the other with a population of 1,000,000 voters, most subjects are more confident in the poll taken in the smaller city. Investigators hold (e.g. Bar-Hillel, 1979, p. 251) that in such a case only absolute sample-size, and not sample-to-population ratio, is related to expected accuracy. So they think that most subjects are mistaken in the importance they attach here to proportionality.

But these subjects need not be mistaken at all. According to the Norm Extraction Method we should choose an interpretation of their responses that fits the recognisable complexity of their problem. First, we should reject any idea that they are trying to evaluate by reference to sample-to-population ratio the very same function (called by Neyman (1937) the co-efficient of confidence) that is correctly evaluated by reference to absolute sample-size. We may then suppose that these subjects are treating the sample-to-population ratio as a straightforward measure of the weight of the evidence, in the sense that, because it measures how big a fraction the actually available evidence (i.e. the actual sample) is of the total possible evidence (i.e. the population), it thereby measures how small a risk is being taken that the sample is biased in some way by its method of selection. So here we have yet another situation in which probability conceived as natural propensity or appropriate intensity of belief might differ from probability conceived in terms of proportions of possible outcomes. The proportion of samples that resemble their population in respect of some particular statistical magnitude (within a given interval of approximation) is determined just by the absolute size of the sample. But the strength with which we may legitimately believe the selected sample to resemble the population must depend also on the degree of care with which we have sought unbiased evidence from the population. One familiar method–the Baconian one–of exercising such care is to select several samples, under appropriately controlled variations of potentially relevant circumstances. Another method is to increase the fraction of the population that is sampled. The latter method is cruder than the former; and less effective if the population is very large. But it is nevertheless a step in the right direction.

Indeed, it is clear that the absolute size of the sample, on the one side, and the sample-to-population ratio, on the other, may be thought of as determining the values of two distinct functions–confidence and weight, respectively–for measuring the reliability of what the sample indicates. Subjects may therefore be expected to respond in different ways, according to whether they operate implicitly with both functions or just with one. In the former case they can take account of both weight and confidence by amalgamating them in a measure of their strength of belief within a Pascalian framework. And in the latter case they can judge either by weight alone, in which case they will be influenced by sample-to-population ratios irrespective of absolute sample-size, or by confidence alone, in which case they will be influenced by absolute sample-size irrespective of sample-to-population ratios. According to the Norm Extraction Method experiments may thus show how many, or what kind of, people respond in one way or the other to different kinds of instructions. They cannot show that most subjects respond mistakenly unless most individual subjects give irredeemably selfcontradictory responses. And such responses turn out to be very rare indeed.

7. THE EXPERIMENTER AS EDUCATIONAL MIDWIFE

Some investigators (e.g. Pitz, 1980, p. 91) have been concerned that experimental results in this field may sometimes be artefacts of task description rather than true indications of judgmental capacity. But such a concern arises out of using the Preconceived Norm Method rather than the Norm Extraction Method. It assumes a norm to which it supposes that the subjects ought to be conforming, and points to the risk that a misleading task-description may cause deviations. The Norm Extraction Method, on the other hand, treats every response as potentially indicative of the norm that the untutored subject is cued to apply by the task-description and its circumstances, so that no such response is more "artificial" than any other.[6]

Moreover, since the Norm Extraction Method encourages investigators to think of themselves as enquiring which norms are applied by subjects in particular kinds of circumstances, it also promotes the study of which circumstances will tend to elicit the application of norms that are already known not to be applied as much as they might be. Perhaps some investigators have been too eager to devise circumstances in which subjects do not apply some computationally useful norm, such as the law that governs the significance of sample-size, and insufficiently

eager to devise circumstances in which subjects' responses come to show recognition of that norm.

Results were achieved by Tversky and Kahneman (1971 and 1974) which undoubtedly demonstrate that in certain circumstances most subjects ignore the significance of sample size. For example, a set of subjects were asked to consider the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know about 50% of all babies are boys. However the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

Most of the subjects judged the probability of obtaining more than 60% boys to be the same in the small and in the large hospital, and Tversky and Kahneman interpret this result as being contrary to sampling theory and as indicating lay ignorance of the law of large numbers. Now, it is in any case arguable, on the Norm Extraction Method, that in these circumstances most subjects are not so much failing in an attempt to measure Pascalian probabilities correctly, as Tversky and Kahneman supposed, as succeeding in an attempt to measure Baconian ones (Cohen, 1979). They are putting a crude, causal gloss on the question and taking it to ask whether the size of a hospital exercises any causal influence on variations in the boy-girl birth ratio. But, whether or not that reinterpretation is accepted, there is no doubt that the tendency to ignore sample-size is not due to lay subjects' being inherently incapable of recognising its significance unless explicitly taught about it. Though Jones and Harris (forthcoming) replicated Tversky and Kahneman's results both in regard to the same task and in regard to a structurally similar one, they also showed that a substantial majority of subjects did appreciate the significance of sample-size in their performance of the Tversky and Kahneman task if these subjects had been tested previously on a rather simple Piagetian task which was designed to elicit awareness of this significance in a practical context. (First 5, and then 50, marbles were dropped haphazardly into a divided box, and in each case subjects were asked for the likely distribution of dropped marbles between the two parts of the box, shown the actual distribution, and asked to explain it.)

In other words subjects can be led to acknowledge and apply the law of large numbers about sample-size by a procedure which suggests that the law was already implicit in their competence as a piece of tacit knowledge. It required a mathematical genius like Bernoulli to formulate this implication of Pascalian probability for the first time. But, now

that has happened, anyone who learns of the law can teach it to others, either by explicit instruction or by a maieutic method such as that of the Piagetian box. The latter method is like the way in which Socrates in Plato's *Meno* brought the mathematically uneducated young boy to a self-conscious awareness of his knowledge that, for any given square, a square on its diagonal is twice its area. This is not a way of acquiring new knowledge *from* experience, but of becoming conscious, at least for a while, of existing knowledge *through* experience. And Jones and Harris's finding suggests that investigators in the field of cognitive rationality might make some useful contributions to the technology of education if they directed their attention more to searching for circumstances that will bring people to apply correctly a particular principle that is already implicit in their cognitive apparatus than to searching for circumstances in which people do not recognise the relevance of the principle.

There is an analogy here with ambiguous drawings like Wittgenstein's duck-rabbit. One time you see such a drawing as a duck, another time as a rabbit, and some people may not be able to see it as either unless appropriately prompted. So too the problem about the boy-girl birth ratios in the two hospitals is subject to a similar gestalt-switch. It may be seen first as a problem about causal connections and Baconian probability, and then later as one about chance distributions and Pascalian relative frequency. As pointed out in Cohen (1979, p. 401) it is pertinent that Piaget and Inhelder (1975) found that the idea of chance developed later in children than the idea of causality.[7] Again, the Phillips and Edwards (1966) problem about predominant chip-colour–see section III above–may be seen first as a problem about Pascalian degree of justified belief, and then, after calibration (Messick and Campos, 1972), as one about *a priori* Pascalian ratios. Perhaps the more sophisticated–or better quantified–conception of such a problem, if given an opening, will tend to prevail over the cruder and less precise one. Perhaps, that is, subjects cued into the former conception are more inclined to find fault with answers given under the latter conception than subjects cued into the latter are inclined to find fault with answers given under the former. But though lay subjects' construals of their tasks may sometimes be unsophisticated, their responses are not therefore incorrect or irrational. Indeed, it would scarcely be a surprising discovery if here and there laymen were found not to achieve unaided anticipation of problems, and techniques for solving them, that form the standard subject matter of elementary courses in statistical theory. Why else are such courses given?

Moreover there are some cases in which the transition to a more sophisticated or statistical conception of probability would not bring any

worthwhile benefit. It is not at all obvious for example (see section IV above), that a careful gambler would always gain by abandoning his "fallacy". And it is certain (Cohen, 1981a, pp. 328f. and 365f.) that the taxi-cab problem is one to which in real life the statistical answer would generally be less appropriate. In this problem subjects were told that in a certain town blue and green cabs operate in a ratio of 85 to 15, respectively, while a witness identifies a cab in a crash as green and the court is told that in the relevant light conditions he can distinguish blue cabs from green ones in 80% of cases. Subjects were then asked: what is the probability (expressed as a percentage) that the cab involved in the accident was blue? If they view this as a question about degree of justified belief subjects may be expected to bear in mind also any other relevant knowledge that they may happen to have. For example, they may well be aware, quite independently, of the possibility that even a less numerous cab-fleet may dominate in a particular area of a town, or because of bad maintenance or poorly trained drivers be particularly liable to accidents, etc. With such facts at the back of their minds, subjects could reasonably regard the relative size of the two cab-fleets as too light-weight a factor to be allowed to modify the strength of a belief that derives from fairly reliable eye-witness testimony. They would then produce what is in fact the prevalent response: the cab in the accident has a 20% probability of being blue. But if anyone views the question as being about a relative frequency–specifically, the relative frequency of correct identifications of a cab-colour in the town as blue–the correct response would be quite different: approximately 59%. This is the only response that Kahneman and Tversky (1972), Lyon and Slovic (1976) and Bar Hillel (1980) take to be correct. But its correctness depends on imposing an interpretation on the question that excludes any background knowledge of a general kind from being relevant. And in a comparable real-life situation such exclusion would not normally be either practicable or desirable, any more than it profits gamblers to forget that coins or dice are sometimes biased and decks or wheels are sometimes rigged.

In outline, therefore, the position so far revealed by the Norm Extraction Method seems to be the following, though much further experimental work no doubt remains to be done. If causal issues are in the offing, lay subjects tend to construe their problem in Baconian terms (Cohen, 1979, p. 397) and to think of probability non-complementationally as running up from nonproof to proof. This tendency is overridden in favour of Pascalian reasoning if subjects are prompted (e.g. by being asked to rate a probability by a betting-quotient or ratio) to think of probability complementationally as running up from disproof to proof. In their Pascalian reasoning subjects most naturally tend to think of probability

as a propensity or as a measure of justified belief and to construe their tasks accordingly. But this tendency is overridden in its turn if subjects are prompted (e.g. by an exercise in calibration or by professional training) to think of probability in statistical terms as an a firiori given ratio of chances or an empirically estimatable relative frequency. And in no case (see also Cohen, 1981a, p. 325) does there seem any good reason to hypothesise that subjects use an intrinsically fallacious heuristic.[8]

Notes

1. The distinction was drawn in explicit terms in Cohen (1977) p. 6, but was overlooked in Cohen (1981a) pp. 319f. and 363.

2. Strictly speaking, if all the relevant knowledge we have about Jones is that he is a university teacher, then on the relative frequency analysis we estimate the probability of a person's surviving to age 70, given that he is a university teacher, while on the personalist account we estimate the probability of a person's surviving to age 70, given the evidence about the relative frequency with which university teachers survive to age 70.

3. The connection between the two analyses is discussed in Mellor (1982).

4. Cf Adler (forthcoming) on how standing principles of conversational understanding may influence subjects to select an understanding of their task that involves an element of empirical judgment, rather than just logical or mathematical computation. The relevance of such principles to the interpretation of subjects' responses was stressed by Cohen (1981a, p. 326) and is now acknowledged by Kahneman and Tversky (1982, pp. 132-135).

5. Other interesting evidence against the prevalence of a representativeness heuristic is discussed by Olson (1976) and Evans and Dusoir (1977). But there is nothing necessarily wrong with a representativeness heuristic so far as this is just a procedure for judging the Baconian probability of a situation's outcome in terms of the situation's similarity in causally relevant respects to situations that are known to have had similar outcomes. That is the right way to assess a Baconian probability (Cohen, 1977), though of course mistakes can be made in practice if irrelevant similarities are taken to be relevant or relevant ones to be irrelevant.

6. I do not wish to suggest, however, that the only type of cueing to which people's judgments of probability are susceptible is in regard to relevant norms. They can be influenced not only in regard to the nature of the function with which they implicitly operate but also in several other ways, such as in regard to the domain to which they apply this function. Consider, for example, the problem discussed by Weaver (1963). The statement of the problem asserts that there are three chests, each containing two drawers, and that one chest has a gold piece in each drawer, another a silver piece in each drawer, and the third a gold piece in one drawer and a silver piece in the other. The instruction is then: "Choose a chest at random, choose one of its drawers at random, and open the drawer. Suppose it contains a gold piece. What is the probability that the other drawer in this same chest also contains a gold piece?" People tend to reply "$\frac{1}{2}$", whereas the propounders of the problem normally insist that the correct answer is "$\frac{2}{3}$". But in fact there are two different probabilities that might be at issue here. One is the probability that a randomly selected drawer contains a gold coin, given that the drawer adjacent to it contains a gold coin, and this probability is certainly $\frac{2}{3}$. The other probability is that a randomly selected chest contains two gold coins, given that it contains at least one gold coin, and this probability is equally certainly $\frac{1}{2}$. The ordinary reader or hearer of the problem is naturally cued to select the latter interpretation by having his attention directed initially towards the selection of a chest (rather than a drawer).

7. In their earlier papers arguing adult ignorance of the law of large numbers (Kahneman and Tversky, 1974; Tversky and Kahneman, 1971 and 1974) Tversky and Kahneman paid no attention to Piaget's evidence for children's tacit knowledge of that law. But they now acknowledge (Kahneman and Tversky, 1982, p. 130-131) the results that may be achieved by the Socratic method in this connection.

8. I am grateful for helpful comments on earlier drafts to Jonathan Adler, Ned Block, Radu Bogdan, Gillian Cohen, Dan Dennett, and Sir Richard Eggleston, and to participants in discussions at Cambridge, Houston, Lancaster and Tulane Universities and in Boston. I am also grateful to the British Academy for a research readership during tenure of which this paper was written.

References

Adler, J. *Abstraction is uncooperative.*

Alker, H. and Herman, G. (1971). Are bayesian decisions artificially intelligent? the effect of task and personality on conservatism in processing information. *Journal of Personality and Social Psychology*, 19:pp. 31–41.

Bar-Hillel, M. (1979). The role of sample size in sample evaluation. *Organizational Behaviour and Human Performance'*, 24:pp. 245–257.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgements. *Acta Psychologica*, 44:pp. 211–33.

Beach, L. (1966). Accuracy and consistency in the revision of subjective probabilities. *IEEE Transactions in Human Factors in Electronics*, HFE-7:pp. 29–37.

Beach, L. and Peterson, C. (1966). Subjective probabilities for unions of events. *Psychonomic Science*, 5:pp. 307–308.

Caroll, J. (1980). Analysing decision behaviour: the magician's audience. In Wallsten, T., editor, *Cognitive Processes in Choice and Decision Behaviour*, pages 69–76. Erlbaum, Hillsdale, New Jersey.

Charme, W. D. and Peterson, C. (1968). Intuitive inference about normally distributed populations. *Journal of Experimental Psychology*, 78:pp. 269–275.

Cohen, J. and Hansel, C. (1957). The nature of decision in gambling: Equivalence of single and compound subjective probabilities. *Acta Psychologica*, 13:pp. 357–370.

Cohen, L. (1979). On the psychology of prediction: whose is the fallacy? *Cognition*, 7:pp. 385–407.

Cohen, L. (1980). Some historical remarks on the baconian conception of probability. *Journal of the History of Ideas*, 41:pp. 219–281.

Cohen, L. (1981a). Can human irrationality be experimentally demonstrated? *Behavioural and Brain Sciences*, 4:pp. 317–370.

Cohen, L. (1981b). Subjective probability and the paradox of the gatecrasher. *Arizona State Law Journal*, 1981:pp. 627–34.

Cohen, L. J. (1977). *The Probable and the Provable.* Clarendon Press.

Donmell, M. and Charme, M. D. (1975). The effect of byesian feedback on learning in an odds estimation task. *Organisational Behaviour and Human Performance*, 14:pp. 305–313.

Ebbesen, E. and Konecni, V. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In Wallsten, T., editor, *Cognitive Processes in Choice and Decision Behaviour*, pages 21–45. Erlbaum.

Evans, J. S. B. and Dusoir, A. (1977). Proportionality and sample size as factors in intuitive statistical judgement. *Acta Psychologica*, 41:pp. 129–137.

Finetti, B. D. (1931). Sul significato soggestivo del probabilità. *Fundamenta Mathematica*, 17:pp. 298–329.

G.F. Pitz, L. D. and Reinhold, H. (1967). Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology*, 21:pp. 381–393.

Hacking, I. (1975). *The Emergence of Probability.* Cambridge University Press, Cambridge.

Jones, C. and Harris, P. (forthcoming). Insigh into the law of large numbers: a comparison of piagetian and judgement theory. *Quarterly Journal of Experimental Psychology*, (in press).

Kahneman, D. and Tversky, A. (12). On the psychology of prediction. *Oregon Research Institute Research Bullletin*, 1972.

Kahneman, D. and Tversky, A. (1974a). On the psychology of prediction. *Psychological Review*, 80:pp. 237–251.

Kahneman, D. and Tversky, A. (1974b). Subjective probability: A judgement of representativeness. In von Holstein, C. S., editor, *The Concept of Probability in Psychological Experiments.* Reidel, Dordrecht.

Kahneman, D. and Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11:pp. 123–141.

Kyburg Jr., H. and Smokler, H. (1964). Editorial introduction. In H.E. Kyburg, J. and Smokler, H., editors, *Studies in Subjective Probability.* New York.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowlede*, pages 91–195. Cambridge University Press, Cambridge.

Lindman, H. and Edwards, W. (1961). Unlearning the gambler's fallacy. *Journal of Experimental Psychology*, page p. 630.

Lyon, D. and Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, 40:pp. 287–98.

Mackie, J. (1973). *Thruth, Probability and Paradox.* Clarendon Press, Oxford.

Mellor, D. (1982). Chance and degrees of belief. In McLaughlin, R., editor, *What? Where? When? Why: Essays on Induction, Space and Time, Explanation*, pages 49–68. Reidell.

Messick, D. and Campos, F. (1972). Training and conservatism in subjective probability. *Journal of Experimental Psychology*, 94:pp. 335–337.

Nagel, E. (1955). Principles of the theory of probability. In O. Neurath, R. C. and Morris, C., editors, *Foundations of the Unity of Science*, volume 1, pages p. 341–422. Chicago University Press, Chicago.

Neyman, J. (1937). Outline of a theory of statistical estimation, based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London*, Series A:pp. 333–380.

Nisbett, R. and Ross, L. (1980). *Human Inference: strategies aand shortcomings of social judgement.* Prentice-Hall, Englewood Cliffs, New Jersey.

Olson, C. (1976). Some apparent violations of the representativeness heuristic judgement. *Journal of Experimental Psychology: Human Perception and Performance*, 2:pp. 599–608.

P. Slovic, B. F. and Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28:pp. 1–39.

Peterson, C. and Charme, W. D. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, 73:pp. 61–65.

Peterson, C., Ulehla, Z., MIller, A., Bourne, L., and Stilson, W. (1965). Internal consistency of subjective probabilities. *Journal of Experimental Psychology*, 70:pp. 526–533.

Phillips, L. and Edwards, W. (1966). Conservativism in a simple probability inference task. *Journal of Experimental Psychology*, 72:pp. 346–354.

Piaget, J. and Inhelder, B. (1975). *The Origin of the Idea of Chance in Children.* Routledge, London.

Pitz, G. (1980). The very guide of life: The use of probabilistic information for making decisions. In Wallsten, T., editor, *Cognitive Processes in Choice and Decision Behaviour*, pages 77–94. Erlbaum, Hillsdale, New Jersey.

Ramsey, F. (1931). *The Foundations of Mathematics.* Routledge and Kegan Paul, London.

Savage, L. (1954). *The Foundations of Statistics.* Wiley, New York.

Savage, L. (1964). The foundations of statistics reconsidered. In Kyburg, H. and Smokler, H., editors, *Studies in Subjective Probability.* John WIley.

Shafer, G. (1978). Non-additive probabilities in the work of bernouolli and lambert. *Archive for History of Exact Sciences*, 19:pp. 309–370.

Swinburne, R. (1964). Fallsifiability in scientific theories. *Mind*, 73:pp. 434–436.

Tversky, A. and Kahneman, D. (1971). The belief in the law of small numbers. *Psychological Bulletin*, 76:pp. 105–110.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185:pp. 1124–1131.

Vlek, C. (1973). Coherence of human judgement in a limitied probabilistic environment. *Organizational Behavior and Human Performance*, 9:pp. 460–481.

Vlek, C. and Heijden, A. V. D. (1967). Subjective likelihood functions and variations in the accuracy of probabilistic informatoin processing. *Psychological Institute Report*, no. E017-67.

Watkins, J. (1958). Confirmable and influential metaphysics. *Mind*, 67:pp. 344–365.

Weaver, W. (1963). *Lady Luck: the Theory of Probability.* Heineman, London.

15

INDUCTIVE LOGIC 1945-1977

I

The seventeenth century saw the beginnings of two powerful and important ways of conceiving one proposition to support another in cases where the truth of the former is formal-logically consistent with the falsehood of the latter. Early in the century Francis Bacon urged the possibility of discovering causal uniformities from tables of presence and absence – lists of circumstances present, and circumstances absent, when the phenomenon under investigation was found. Bacon thought that natural laws which were so discovered would become more and more certain as they increased in comprehensiveness and subsumed a greater and greater variety of known uniformities, provided that these laws also lead us to new knowledge. A little later Pascal and Fermat laid down principles for a mathematical calculus of chance that Leibniz and Bernoulli interpreted as binding normal judgements of probability – in the law courts, for example, as well as in games of chance.

For a couple of centuries these two systems of ideas tended to be considered apart from one another.

Like a good Baconian Leibniz considered one scientific hypothesis to be more probable than another if it was simpler, explained a larger number of phenomena from a smaller number of postulates, and allowed the prediction of as yet unknown phenomena.[1] But, apparently, he never tried to represent or elucidate these criteria within the calculus of chance, despite his application of the calculus to other kinds of probability judgment. Hume actually went on to distinguish two different types of probability when he argued that

> probability or reasoning from conjecture may be divided into two kinds, *viz.* that which is founded on *chance*, and that which arises from *causes*.[2]

William Whewell, J. F. W. Herschel and J. S. Mill reformulated Bacon's ideas in the light of post-Baconian scientific developments, but never attributed a Pascalian structure to judgements about how much this or that variety of evidential data supports a stated universal generalisation about causal processes.

Equally, when mathematicians refined and developed the theory of Pascalian probability, they pointed out its value *within* science, in calculations aimed at drawing out the best value for a quantitative parameter from a multitude of varying measurements.[3] Or they argued that an empirical hypothesis increased its probability with the mere number of observed instances irrespective of their variety.[4] But they did not normally propose applying Pascalian principles to the main logical problem *about* science with which the Baconian tradition had been grappling – viz. the problem of how to assess the strength of the support that is given by experimental variations of evidential circumstances to a universal hypothesis about causal connections.

Nevertheless, in the latter half of the nineteenth century, as the mathematics of Pascalian probability became better and better understood, while the logical structure of Baconian induction remained wholly unformalised, the temptation to treat the former as a theory of the latter became more and more obviously attractive. W. S. Jevons certainly succumbed to this temptation;[5] and in the 1920's J. M. Keynes[6] in England, and Jean Nicod[7] in France, found it quite natural to assume without question that judgements of inductive support from variations of evidential circumstance should conform to Pascalian principles. A great part of the more recent history of the subject has also consisted of work done on the basis of this unquestioned assumption. But, as we shall see, the assumption has turned out to be a rather frustrating one, and its rejection permits a breakthrough into some rewardingly new areas of logico-mathematical analysis.

II

It would be a mistake to suppose, however, that the only important constraints on recent development in inductive logic have been set by Pascalian ideas about probability, on the one side, and scientists' norms of evidential assessment, on the other. In addition to those two factors, certain well-known puzzles or paradoxes have made a major contribution towards determining the direction that developments have taken. These paradoxes have been as important for inductive logic as have other paradoxes, like the Liar and Russell's antinomy, for deductive logic: I shall mention just three of them.

In 1945 C. G. Hempel pointed out,[8] in effect, that three apparently plausible assumptions about the concept of confirmation are not cotenable. These assumptions are as follows: –

1 If the antecedent and consequent of a generalized conditional are conjointly satisfied, this confirms the generalisation to at least some extent. For example, an object that is both a raven and black confirms the generalization 'Anything, if it is a raven, is black.'

2 Logically equivalent propositions are equally confirmed by the same evidence. So evidence that confirms 'Anything, if it is a raven, is black' must also confirm its contrapositive equivalent 'Anthing, if it is not black, is not a raven'.

3 A typical object that is neither black nor a raven, like a white handkerchief, does not confirm 'Anything, if it is a raven, is black' (despite the fact that it confirms the contrapositive equivalent of this generalisation).

Obviously one or other of these assumptions has to be modified and Hempel's own proposal was to replace (1). Another proposal[9] has been to sidestep (2), by saying that evidence which satisfies a generalisation like 'Anything, if it is a raven, is black', will selectively confirm the generalisation only if it also disconfirms the generalisation's contrary, viz. 'Anything, if it is a raven, is not black.' Yet another proposal[10] has been to reject (3), on the ground that, because so many more things are not ravens than are ravens, the finding of some non-black things that is not a raven is merely of very low, not null, confirmatory value. But it is unprofitable to discuss the merits of any one such solution of this paradox except as part of a general theory of inductive reasoning which is also competent to solve all other relevant paradoxes also.

In 1953 N. Goodman propounded[11] another paradox about the same criterion of confirmation. Suppose that all emeralds examined before a certain time t are green. Then at time t, according to this criterion, all our past observations of emeralds confirm the hypothesis that all emeralds are green. But the same observations also confirm the hypothesis that all emeralds are grue, where the novel predicate 'grue' is defined as applying to all things examined before t just in case they are green and to other things just in case they are blue. Hence our observations appear to confirm equally well, two inconsistent predictions: the prediction that all emeralds subsequently examined will be green and the prediction that all emeralds subsequently examined will be blue.

Goodman's own solution of his paradox was to supplement Nicod's criterion by a linguistic requirement. He proposed to evaluate rival hypotheses in accordance with their 'degree of projectibility', where one hypothesis is said to be more projectible than another if the predicates in terms of which it is formulated – or predicates co-extensive with them – have occurred more often in successfully predictive hypotheses. (Also, if a predicate P applies only to certain mutually disjoint classes, and Q applies to one of these classes, then P, or any predicate co-extensive with P, is said to be as good as Q for these purposes.) If two inconsistent hypotheses were equally well instantiated by our observations, the more projectible of the two was to be regarded as the better confirmed. But it is arguable that, if the founders of modern chemistry had behaved as Goodman suggests, we should still be talking about earth, air, fire and water, instead of about hydrogen, lithium, beryllium, boron, etc.[12]

A third paradox of fundamental importance to the subject was formulated in 1961 by H. E. Kyburg.[13] This paradox is not concerned with the relation between evidence and hypothesis, but with how one may treat a hypothesis once it has achieved a satisfactory degree of evidential confirmation. Again three apparently plausible assumptions are not co-tenable: –

1 It is reasonable to accept any hypothesis H that, after thorough enquiry, retains a Pascalian probability (on the available evidence) of $1 - \epsilon$ (where ϵ is as small a positive number as you care to make it).

2 It is reasonable to accept any logical consequence of a set of hypotheses that it is reasonable to accept.

3 It is not reasonable to accept an inconsistent set of hypotheses.

These assumptions are not co-tenable in relation to, say, a lottery that sells a million tickets, only one of which can win the prize. In the case of each of the million tickets it is reasonable to accept the hypothesis that that ticket is not the winning one. But this conjunction of hypotheses is inconsistent with another hypothesis, which it is also reasonable to accept, viz. the hypothesis that just one ticket will win. Kyburg's own solution of the paradox was to restrict (2), by confining its application to logical consequences of individual hypotheses, as determined by some original partition of possibilities. Other philosophers have modified (1) in various ways, and some have even been prepared to reject (3). But no solution can be satisfactory unless it has a rationale that is independent of its ability to resolve this particular paradox. *Ad hoc* solutions, which are justifiable only by their possession of that ability, cannot provide us

with adequate intellectual compensation for having to give up the conjunction of three such plausible assumptions.

III

R. Carnap[14] was the first philosopher in the post-1945 period to undertake a large-scale systematic investigation of inductive reasoning. Suppose a language to have one or more families of mutually incompatible primitive predicates (e.g. the family 'red', 'blue', white', etc.). Then a state-description, for a given domain of objects, may be defined as a conjunction of propositions that ascribe to each individual object exactly one primitive predicate of each family. A weight is to be assigned to each state description so that all the weights sum to unity. Different ways of thus assigning weights to state-descriptions are conceivable: one such would be to assign each the same weight. The range of a proposition p is defined as the set of all the state descriptions in which p occurs. Once a measure m has been defined for these ranges, as the sum of the weights assigned to the various state-descriptions in the range, it is possible to define the degree of confirmation that one proposition e gives another h, which is written $c(h, e)$, as the ratio of $m(h\&e)$ to $m(e)$. The intuitive idea is that e's support for h is measured by how far the various possible states of affairs in which both h and e hold good overlap with the various possible states of affairs in which e holds good. And, since range-measures of this kind conform to the principles of Pascalian probability, the resultant confirmation-functions do also, in virtue of Bayes' law. Out of a continuum of desirable range-measures Carnap preferred those producing a higher and higher degree of confirmation for the hypothesis that an object has a certain property according as the evidence comes to include more and more objects that already have the property. Indeed, as J. Humburg and H. Gaifman[15] later showed, it can be proved in Carnap's final version of his system (with the help of a plausible additional axiom) that the confirmation of a singular statement h_1 is always increased when, in addition to original evidence stated in e, which is consistent with h_1, we accept a new evidential statement h_2 which reports another instance of the same property as h_1 reports.

But the discovery of a proof for this principle, though welcomed by Carnap, was in fact a heavy blow to his claim to be explicating natural-scientific modes of discourse about inductive support. The discovery showed that this 'principle of instantial relevance', as Carnap called it, is rather deeprooted in his confirmation-theory. And the harsh truth is that functions for which the principle is provable are cut off therewith from granting the support that reports of successful experimen-

tal tests are supposed to give to the hypotheses tested. The reason is that e might already report an instance of the property described by h_1 and h_2 cannot be regarded as stating evidence that increases the support for h_1 merely because it describes yet another instance of the same property. Admittedly a test-result has to be replicable in order to count as genuine evidence in favour of the hypothesis concerned. But the test-result's actual replication (perhaps in another laboratory at a later date) merely confirms its replicability. To get stronger confirmation for the hypothesis itself one would have to construct a different experiment, either testing a different consequence of the hypothesis, or testing the same consequence more thoroughly – i.e. with better controls and greater safeguards against the interference of factors not mentioned in the hypothesis. So, because Carnap's type of confirmation theory requires $c(h_1, h_2\&e) > c(h_1, e)$ in the circumstances described above, it cannot avoid being closer to what Bacon contemptuously described as induction by simple enumeration than to the induction by variation of circumstances that Bacon, Whewell, Herschel and Mill rightly took to be characteristic of scientific reasoning about the probative force of experimental results.

Carnap's inductive logic has often been criticised for its failure to assign plausible values to confirmation-functions where the hypothesis to be confirmed is a universal generalisation. If the hypothesis is a singular proposition or conjunction of propositions the logic allows a wide spread of possible values, in the closed interval $\{0, 1\}$, to the function $c(h, e)$, so as to vary with the actual relation of e to h. But, if h is a universal proposition, its degree of confirmation from favourable instances gets very small as its domain increases, and for domains of infinite size its value is always zero. Since some domains of scientific enquiry may be infinite, this seems a flaw in Carnap's logic. But Carnap himself did not think that it was one. Carnap argued that, for a scientist or engineer to construct his predictions on the rational basis of past evidence, it was not necessary for him to route his inferences via appropriate generalisations. Indeed, while a man who had seen many white swans and no non-white swans might be willing to bet on the next swan's being white, he might well be unwilling to bet that every swan in the universe is white.

Carnap's attitude here was curiously the inverse of Karl Popper's. Popper denied that science had any need for inductive logic, and claimed that scientists sought instead to discover the boldest and simplest hypotheses that resisted falsification. Thus while high prior probability could help a hypothesis to achieve a high degree of confirmation by Carnapian criteria, it was high prior improbability that helped a hypothesis to qualify for scientific esteem according to Popper's corroboration-

functions.[16] Correspondingly each philosophy, as a logic of science, had both a merit and a defect that the other philosophy lacked. Carnap could claim to represent a certain way of measuring evidential support for the singular predictions that are essential to technology: an engineer must know how reliable, on the evidence of past experience, will be the bridge that he is actually building. But Carnap could not represent a plausible way of measuring evidential support for the highly general propositions that are needed for theoretical explanation and systembuilding in natural science, like Newton's laws of motion. Popper, on the other hand, could offer a plausible criterion for choosing between one system of scientific theory and another, if the only thing at issue was explanatory power. But he could not provide any adequate guidance for the technologist on questions of reliability. Hypothesis h_2 may be so much bolder than h_1 that we cannot yet test it as thoroughly as h_1 has been tested; and then, even if h_2, like h_1, has not yet been refuted, we can scarcely regard it as being more reliable than h_1 merely because it is bolder. Popper claimed that his falsificationist methodology had solved Hume's problem about induction: there was no need to try and *prove* propositions about the future. But he could only maintain this claim by systematically ignoring the intellectual needs of engineers, navigators, agriculturalists and other technologists. Again, Carnap's adoption of a range-measure ratio, to evaluate degree of confirmation, rendered his inductive logic relatively immune to Hempel's paradox and his prior choice of language protected him, albeit artificially, from Goodman's. But an acceptance-criterion constructed in terms of one of his confirmation-functions would certainly be hit by Kyburg's paradox. On the other hand, while a defender of Popper's views about the logic of science has to work very hard indeed to show that they are not hit by analogues of Goodman's paradoxes, Popper's exposure to Kyburg's paradox is substantially reduced by the fact that he seeks high *im*probability and seeks it for general theories.

IV

J. Hintikka contributed an important new idea[17] to range-based inductive logic in 1964. He reviewed the difficulty about getting adequate confirmation for universal propositions in Carnap's system, and traced it to the way in which the measure of a proposition's range comes to depend on the number of individuals in the domain.

Carnap measured the range of a proposition by the sum of the weights of the different state-descriptions in which the proposition held good. He preferred not to assign equal weights to every state description, however, because this led to a confirmation-function that was indifferent to the ac-

cumulation of favourable evidence. Instead he proposed assigning equal weights to every structure-description, where each structure-description is a disjunction of all the different state descriptions that can be transformed into each other by permuting the names of individuals. Suppose, for example, that the language consists of a family of just two primitive monadic predicates, and just two names of individuals, plus the usual logical connectives: in such a language four state-descriptions can be constructed but only three structure-descriptions. Carnap's idea was then to divide the weight which a structure-description receives into equal parts, one for each of the state-descriptions that it disjoins.

But this still makes the measure of a universal proposition's range depend in part on the number of individuals in the domain, and as Hintikka remarked, it is rather perverse to start one's inductive logic by assuming that one is already sufficiently familiar with the whole of one's universe to know its cardinality. Accordingly Hintikka proposed a type of range-measure that was, in the case of universal propositions, independent of this number.

Suppose that we are given k primitive monadic predicates. By means of these predicates and propositional connectives we can partition our domain, whatever its size, into 2^k different kinds (empty or non-empty), distinguished by the various complex predicates that describe them. Then we can describe a possible world by a sentence describing all the non-empty kinds of individuals in it. Such a sentence, which Hintikka calls a 'constituent', will be a consistent conjunction of existentially quantified complex predicates, conjoined with a universal quantification of their disjunction. An equal weight may be given to each constituent (or kind of possible world), and the weight of each constituent may be divided among all the state-descriptions that make it true. But, since every consistent universally quantified sentence has an equivalent (its distributive normal form) that is a disjunction of constituents with the same predicates, the measure of a universal proposition's range is independent of the number of state-descriptions that make any particular constituent true and thus quite independent of the numbers of individuals of various kinds in the universe or of the total number of individuals. Hintikka also showed how his system could in principle cover sentences with polyadic predicates or more than one quantifier. His confirmation-functions, like Carnap's conformed to the principles of Pascalian probability. But they nevertheless resembled Popper's corroboration-functions in allegedly recommending to us the simplest generalisation compatible with the evidence because, for Hintikka, the simplicity of a generalisation varied inversely with the structural complexity of its distributive normal form, rather than with its probability.

Hintikka's inductive logic is thus a definite improvement on Carnap's. But it still suffers from a number of serious disadvantages.

(1). As we have already seen, neither type of logic captures the characteristic scientific attitude towards the bare replication of existing evidence – viz. that it confirms the legitimacy of the evidence but not the hypothesis itself.

(2). Neither type of logic can in practice yet go any distance towards measuring degrees of confirmation for hypotheses that contain polyadic predicates, because of the difficult problems that arise in combinatorial arithmetic. Neither type of logic, therefore, is yet able to provide a reconstruction of scientific inference beyond a very elementary level. Treatment[18] of the role of theoretical concepts in scientific inference can hardly be very convincing if the language of analysis is confined to first-order monadic predicates. Hintikka's confirmation-functions are also subject to those restraints on computability that derive from the non-existence of a decisionprocedure for the logic of polyadic predicates.

(3). Hintikka's inductive logic seems at first sight to be better than Carnap's at avoiding Kyburg's paradox. But in fact it is not. Hintikka and Hilpinen[19] seek to avoid the trouble that Carnap's theory encounters here by treating the fundamental unit of acceptance as a universal proposition, not a singular one. No analogue of the lottery paradox can then be constructed. But a singular proposition is then said to be acceptable if and only if it is a substitution instance of an acceptable generalisation – as if one ought not to accept that it will rain to-morrow unless one accepts that it will rain for all eternity.

4. Perhaps the most serious flaw is this. Both Carnap's and Hintikka's logic assume a pre-existing language, and the confirmation-assessments that they derive therefrom are accordingly analytic and *a priori* in character (as also are Popper's assessments of the degree to which one proposition corroborates another). Consequently all expressions of a given linguistic category have to be put on a level with one another. At some point equipartitions of weight are assumed – between all state-descriptions, or all structure-descriptions, or all constituents – as if all circumstances in nature are equally important (at the appropriate level of complexity). Hintikka's logic is sensitive to variations of evidential circumstances, but it treats all such variations alike, at any particular level of complexity. Yet in the actual conduct of scientific reasoning some evidential circumstances (i.e. some families of predicates)

may be deemed to be much more important than others in relation to hypotheses of a particular kind. A patient's previous medical history turns out to be more important than the colour of his eyes in relation to most hypotheses about the safety of a drug for medical purposes. Not surprisingly, therefore, most scientists take assessments of how well a certain experiment's results support a given hypothesis to be empirically corrigible. When pharmacologists learned that some drugs like thalidomide had teratogenic properties, they learned thereby to incorporate trials on pregnant rats in any experimental test on drugs of that kind. At the same time they had to revise any assessments of the support given by previous test-results to hypotheses about the non-toxicity of this or that drug. So, if we want to represent inductive assessments in natural science by a two-place function $c(h, e)$ that maps ordered pairs of propositions on to numbers, we ought not to treat values for the function as emerging analytically from some independently existing sub-language. The fashioning of the relevant sub-language is an indispensable part of the process of scientific reasoning. Both Bacon and Whewell implicitly recognised this long ago, when they emphasised the significance, for inductive purposes, of the invention of new concepts. It is quite understandable that some modem logicians should have hoped an inductive logic might issue in analytic, *a priori* truths which would be comparable with the analytic, *a priori* truths of standard deductive logic. Particular equations of the form $c(h, e) = r$ were to be comparable with particular instantiations of Russellian theorems. But, though that kind of inductive logic may conceivably have a useful application elsewhere, it certainly misses essential features of inductive reasoning in experimental science. If there is any parallelism between deductive logic, on the one hand, and the inductive logic of experimental reasoning, on the other, it has to be sought along another dimension.

V

I. Levi has proposed[20] a rather different approach to the problems of inductive logic. He too supposes, like Popper and Carnap and Hintikka, that a sublanguage is given which determines the kinds of hypothesis and evidential report that can be considered. But he considers the inductive reasoner as a man who ought rationally to choose that option which bears maximum expected utility, and the utilities involved are understood to be epistemic ones. These cognitive options are ordered by two rules. Correct answers ought to be epistemically preferred to errors; and correct answers, or errors, that afford a high degree of relief from agnosticism ought to be preferred to correct answers, or errors, respectively, that afford a low degree of relief from agnosticism. In order to determine

the degree of relief from agnosticism that a given sentence affords, Levi constructs a measure of a sentence's content that is based on the ultimate partition of the sub-language into available hypotheses. It then becomes possible to determine, relative to a given statement of evidence, the cognitively optimific choice of hypothesis.

Levi's content-measures are Pascalian improbabilities, and his system therefore has obvious affinities to Popper's theory of corroboration as a reconstruction of the intellectual basis for acceptance. But Levi can justifiably claim that his own decision-theoretic system establishes a much closer rational connection than Popper or Hintikka does between his proposed criteria for evidentially-warranted acceptance, on the one hand, and certain plausibly supposed ends of scientific enquiry, on the other. He can explain why people should want hypotheses that score well by his criteria. Also, Levi's system is as immune as Carnap's or Hintikka's to the paradoxes of Hempel and Goodman, because it has replaced their simple, qualitative criterion of confirmation for hypotheses in every day language by an appropriately sophisticated quantitative criterion for hypotheses in a rather restricted artificial language. But his resolution of Kyburg's paradox seems to be entirely *ad hoc*. Levi proposes to limit the application of his acceptance rule to propositions composing an ultimate partition of the domain of assessment, such as 'Ticket 1 will win' and 'Ticket 1 will not win'. In this way no conjunction of propositions from different ultimate partitions will be eligible for assessment and the lottery paradox cannot be constructed. But the trouble is that Levi's restriction seems to have no other motivation or rationale than its capacity to avoid difficulties about a possible lottery. If it were not for those difficulties, we might well feel that all conjunctions of propositions should be just as eligible for assessment as their conjuncts. In the complexities of practical life we need constantly to combine and recombine beliefs about different issues.

Levi's system has the same overall defect of apriorism that taints Popper's ideas about corroboration-functions, and vitiates the Carnap-Hintikka theory as a logic of experimental reasoning. Inductive assessments are derived from a postulated sub-language but nothing is said about the extralogical terms of this language except that there will be 'as many ... as the occasion demands'.[21] So, in a period (1950-1970) when inductive logicians gave such cavalier treatment to the actual facts of scientific reasoning and typically justified their theories by appealing to their own or their readers' 'intuitions',[22] it is scarcely surprising that many philosophers came to be rather sceptical about the extent to which logical analysis can shed any light on the nature of science. Indeed, if Popper, Carnap, Hintikka, Levi and others had produced more convinc-

ing analyses, the anti-analytical reaction of philosophers like Hanson, Kuhn, Feyerabend and Toulmin would perhaps hardly have got off the ground. But, as has so often happened before in human intellectual history, the reaction was overdone. The limitations inherent in a certain onesided approach to a subject were attributed to the subject itself. Instead of finding fault just with apriorist versions of inductive logic, these critics supposed no logic at all to be capable of representing actual scientific reasoning.

VI

M. Hesse's account of scientific inference[23] avoids the fallacy of apriorism. She interprets the (Pascalian) probability of a hypothesis on given evidence in personalist terms; so that in the long run, when all evidence that is directly or indirectly relevant has been taken into account, different scientists' probability-assessments may reasonably be expected to coincide, if each scientist treats all his assessments at any one time as if they were the elements of a coherent betting strategy. One proposition e is then said to confirm another h, if positively relevant to it, i.e. if $p(h, e) > p(h)$; and the extent of such confirmation may be measured by the extent to which $p(h, e)$ exceeds $p(h)$. In this way degree of confirmation becomes sensitive to the changing state of scientists' information, rather than just to the structure of some arbitrarily assumed sub-language. Of course, a personalist interpretation of Pascalian probability-functions ought not to be considered available where universal propositions over an unbounded domain are taken as possible arguments for the function. A reasonable man does not envisage betting on the truth of such a universal proposition, since a single observable event might demonstrate that he had lost his bet but no determinate sequence of observable events could ever demonstrate that he had won it. However, Hesse, like Carnap but unlike Popper or Hintikka, does not envisage a scientific hypothesis that acquires inductive confirmation as being a universal proposition over an unbounded domain. Instead she thinks of it as having some such form as 'All the next n P's in a limited region of space and time are Q.' Consequently, like Carnap also, Hesse argues that enumerative induction is more fundamental than induction by variation of circumstance. But she does make rather more allowance for the latter than Carnap does. First, she emphasises that, where a finite list of hypotheses is assumed, the elimination of some hypotheses makes a disjunction of the others more probable. Secondly, she points out that, where there is a substantial analogy between the phenomenon described by one hypothesis and the phenomenon described by another, each increases the probability of the other.

Hesse's choice of restricted generalisations, rather than truly universal propositions, as candidates for confirmation, is in fact motivated by her adoption of positive Pascalian relevance as the criterion of confirmation – even though her personalist interpretation of probability-functions also compels this choice. The point is that her positive relevance criterion compels her to accept the so-called converse entailment condition. Where one hypothesis, h, entails another, g, and we have both $1 > p(h) > 0$ and $1 > p(g) > 0$, Bayes's theorem gives her $p(h, g) = [p(h)/p(g)] > p(h)$, and so g confirms h. But an intolerable paradox results if anyone accepts both this converse entailment condition and also the so-called special consequence condition, viz. if f confirms h and h entails g, then f confirms g. For, by accepting both conditions, it is possible immediately to prove that any two propositions confirm one another: if f & g is logically equivalent to h, then by the converse entailment condition f confirms h and so, by the special consequence condition f confirms g. Accordingly Hesse rejects the special consequence condition. Yet that condition has a good deal of plausibility in its application to general scientific theories. For example, one might well suppose that the confirmation afforded to Newton's theory of gravitation by the Keplerian (or quasi-Keplerian) laws of planetary orbits and the Galilean (or quasi-Galilean) law of falling bodies flows down to that theory's predictions about the orbits of comets or of space-satellites. In order to avoid countenancing such a source of plausibility for the special consequence condition Hesse proposes to reject theories like Newton's as candidates for confirmation and to treat them rather as rules for analytical inference from one kind of restricted generalisation to another. The inductive strength of the theory will then presumably be measured by the increase of probability that one such generalisation can give to another.

Here at last we have a confirmation-measure that is not only sensitive to changes in scientific opinion and to differences in the degree of relevance between one piece of evidence and another, but also tries to grapple realistically with the part played by theoretical systematisation in inductive reasoning. This latter factor was emphasised long ago by Bacon; and Whewell, calling it the 'consilience of inductions', was able to furnish a number of impressive illustrations of it from the history of science between Bacon's day and his own. But it was never treated adequately in the work of Carnap, Hintikka or Levi. Nevertheless Hesse's inductive logic still suffers from at least three major implausibilities as an analysis of actual scientific reasoning.

One of these is the logic's refusal to treat general theories like Newton's as propositions that can be either confirmed or disconfirmed. It

is natural to suppose, as Bacon did, a kind of spectrum in people's hypotheses about nature, beginning at one extreme with hypotheses about obvious, large-scale phenomena in a man's immediate spatio-temporal environment, and going on, through more and more extensive speculation, to hypotheses about laws that control the fundamental constituents of reality throughout the universe. Inductive logic, we might suppose, can deal more easily with hypotheses of the more limited type of scope. But its domain can scarcely be restricted to these. Nor is this conception just a prejudice of philosophers or logicians. Hesse invites us to accept a kind of discontinuity in the pattern of scientific reasoning that has no warrant whatever in the ways in which any leading chemists or physicists have expressed themselves.

A second difficulty about Hesse's inductive logic arises from her use of a positive relevance criterion that is formulated in terms of Pascalian probability. Where, say, $p(g) = p(h) = 0.2$, and $p(g, e) = 0.9$, and $p(h, e) = 0.1$, we shall have to hold, according to Hesse's criterion, that e confirms g and disconfirms h. Yet, if g and h here are independent of one another, e will confirm the conjunction of g and h. Hence, so far as our inductive logic is to afford a basis for our technological judgements, we are told here that there is confirmation for exploiting the conjunction of g and h in constructing our bridges or airplanes, despite the fact that there is disconfirmation for h on its own. It is difficult to think that an inductive logic which issues in such judgements could ever be much used in practice.

A third difficulty is that though Hesse's logic is not hit by Hempel's or Goodman's paradox it has no way of meeting the challenge of Kyburg's, unless perhaps in the same *ad hoc* way as that in which Levi tried to meet it. In fact Hesse herself prefers to suppose that inductive logic can afford no basis for a criterion of rational acceptance. But to suppose this is to sacrifice one of the major purposes that might be thought to make investigation of the subject worthwhile.

VII

All the above-mentioned post-1945 logicians have shared one assumption. They have all assumed that any function which assesses evidential support must either itself conform to the principles of Pascalian probability, like Carnap's and Hintikka's confirmation-functions, or must at least be built up from functions which so conform, like Popper's corroboration-function or Hesse's confirmation-measure. Yet this was an assumption which writers in the classical, Baconian tradition did not share, as we have already seen. Correspondingly these recent writers have envisaged the aim of inductive reasoning in natural science as some

kind of maximisation in relation to truth. The aim is either to know as much of the truth as possible, or to be as near as possible to the true answer for a certain question, or to have as large a chance as possible that one has got hold of that answer, or at any rate to have a larger chance than one would have had without the evidence. But again, the aim of classical inductive logic was a different one. It sought the discovery of laws, not just of truths. It was interested in what causes what, not just in what happens. Bacon's terms for such a law was a 'form', and knowledge of the form of heat, on his view, would enable a man to superinduce heat under suitable conditions. The propositions stating such laws he called axioms, and the higher a scientist mounts on the ladder of axioms the greater the certainty he could have. And all this was to be achieved by studying tables of presence and absence.

At the heart of Bacon's inductive logic, therefore, lies a marriage between the idea of induction by variation of the evidence, traces of which are to be found earlier in the writings of Robert Grosseteste, William of Ockham and others, and the idea of intellectual progress through a hierarchy of forms, which we owe to Plato. Correspondingly, if we wish to develop and systematise Bacon's logic, our theory[24] will need to have two main components. First, it needs to embrace what I shall call 'the method of relevant variables', as a representation of the way in which, in any particular field of enquiry, the reliability of a hypothesis may be graded in the light of appropriately varied experimental evidence. Secondly, it needs to embrace a generalised modal logic which will represent those grades of inductively attestable reliability as steps on a ladder that mounts towards the statement of a law.

In experimental science the reliability of a hypothesis is identified with its capacity to resist falsification under cumulatively more and more challenging conditions. So, in relation to each given category of scientific hypotheses, it is necessary to form a higher-order empirical hypothesis about the variables – non-exhaustive sets of mutually exclusive circumstance-types – that are inductively relevant to them, where a variable is said to be relevant if each circumstance-type (or 'variant') in that variable has falsified at least one generalisation of the category in question. This list of relevant variables must also be ordered, in accordance with the supposed falsificatory potential of the various variables. The list is then extended by prefacing that variable which contains all the circumstances mentionable by the antecedents of generalisations in the category. A scientific hypothesis may be supposed to be tested with first grade thoroughness when every variant of the first variable in the final list is present in turn, so far as this is possible, but no variant of any other relevant variable is present. A hypothesis is tested with sec-

ond grade thoroughness when each possible combination of a variant of the second variable with a variant of the first (and no variant of any other relevant variable) is present in turn. And so on, with cumulatively more and more complex tests. Accordingly at least first-grade inductive reliability may be inferred for a generalisation from a proposition reporting, in effect, that it has passed a test of first-grade thoroughness, and higher grades of reliability may be inferred from propositions reporting successful test-results of correspondingly higher grades. But failure to pass, say, a third-grade test implies that a generalisation has at most second-grade reliability.

The replicabiiity of test-results is assumed. So that if contradictory gradings are inferable from true evidential reports, we need to revise either the list of relevant variables (as in the thalidomide case) or the terminology from which hypotheses are constructed (as when new scientific concepts are introduced). The discovery of such a contradiction is not a deep blow to the whole system of inductive logic, as it would be for an apriorist theory like Carnap's or Hintikka's, but represents a fruitful stimulus to scientific imagination and research.

Moreover this mode of grading inductive reliability – by the method of relevant variables – can be applied *mutatis mutandis* to causal generalisations, to correlations between quantitative variables, and to comprehensive scientific theories, as well as to elementary generalisations about qualitative characteristics. It produces a very straightforward analysis of consilience. And it can also be applied to appropriately modified versions of hypotheses, where the unmodified versions meet unfavourable test results. Thus if a generalisation is falsified by certain variants of a relevant variable but not others, a mention of the latter can be introduced into the antecedent of the generalisation as so to exclude falsification by the former. In this way a higher grade of inductive reliability can be maintained for a hypothesis, though at the cost of accepting a lower grade of simplicity for it. Nor is the method of relevant variables restricted in any way to monadic predicates. But, unlike Carnap's, Hintikka's, Levi's or Hesse's system, the method of relevant variables does not measure anything. It merely ranks inductive support, by mapping ordered pairs of propositions, of a particular category, on to the first n integers ($n \geq 0$) where there are $n - 1$ relevant variables. Consequently the method can have no pretensions to be able to compare the grades of support that two hypotheses have if they belong to quite different fields of scientific enquiry, except in the limiting cases of no support or full support. On the other hand the method does not apply only to the natural sciences. It applies equally well to other fields of inductive enquiry, such as to ju-

risprudences that rely on precedent and derive rules of law from previous judicial decisions.

If inductive reliability is graded in this way, the principles that govern compatibilities and incompatibilities between such gradings are demonstrably incapable of being mapped on to the Pascalian calculus of probabilities. They are representable instead within a generalisation of the C.I. Lewis modal logic S4. For example, the so-called special consequence condition emerges, as with confirmation-functions that determine Pascalian probabilities. We also get a similar uniformity principle, or principle of symmetry, whereby the inductive reliability of a singular proposition is invariant under any unform transformations of its individual constants. But the logic has certain importantly non-Pascalian features. A hypothesis can have greater than zero reliability even when the evidence includes some counter-examples to it; and the conjunction of two hypotheses has the same grade of reliability as either of its conjuncts, if they are equally reliable, or as the less reliable of the two if they are unequal in reliability. Also, mere multiplicity of instances – repetition of precisely the same test-result – neither increases nor decreases inductive support for a generalisation.

This kind of inductive logic supplies both a dyadic function $s[H, E]$ that grades the support given by E to H, and a monadic one $s[H]$ that grades the natural reliability of H – i.e. the strength of its resistance to falsification by relevant variables. But in addition to dyadic and monadic support-functions for propositions of a given category, the logic also generates dyadic and monadic probability-functions of a characteristically non-Pascalian type, viz. $p_I[S, R]$ and $p_I[S]$. Any first order generalisation of the form $(x_1)(x_2)\ldots(x_n)(R \rightarrow S)$ – where x_1, x_2, ...and x_n are all the individual letters free in R and S – provides us with a rule of inference from the satisfaction of its antecedent to the satisfaction of its consequent, and the strength of such a rule may be graded by the inductive reliability of the generalisation. Hence the inferability of the consequent from the antecedent may be regarded as a form of probability that is graded by the reliability of the covering generalisation. For example, $p_I[Sa, Ra] = s[(x)(Rx \rightarrow Sx)]$.

Moreover, just as the inductive reliability of a generalisation may be raised by introducing appropriate modifications into its antecedent, so too the inductive probability of Sa on Ra & Va may be higher than on Ra alone. Inductive probability-functions thus capture the principles behind that kind of everyday probability judgement in which the probability of a conclusion rises as the weight of evidence – the extent of relevant information – increases, where the balance of the evidence in any case favours the conclusion. These functions are therefore par-

ticularly well suited to the task of providing a foundation for rules of rational acceptance. After all, there are very many kinds of issues on which it is both folly to make up one's mind until one has enough information, and also impracticable to wait for certainty. We need, as it were, a high but non-maximal grade of inductive probability on available evidence. At the same time inductive probability-functions can afford a basis for rational acceptance, even in relation to predictions about one's not winning a lottery. This is because by inductive standards one needs considerably more information than the number of tickets in order to be justified in believing that one will not win: for example, information that the draw is likely to be rigged in favour of the organiser's nephew might suffice. So Kyburg's paradox does not arise here.

The logical syntax of inductive probability judgements derives directly from that of inductive reliability, but the dyadic probability-function does not behave in quite the same way as the dyadic support-function. For example, though $p_I[Sa, Ra]$ conforms to the same conjunction principle as $s[H, E]$, it does have zero-value in normal cases in which Ra contradicts Sa. But a monadic probability-function may be defined that makes $p_I[Sa]$ equal to $p_I[Sa, Sa \vee -Sa]$. It then turns out to be demonstrable that a monadic inductive probability-function always has precisely the same value, for a given argument, as a monadic inductive support-function. I.e., $p_I[Sa] = s[Sa]$. Note also that the prior inductive probability of a proposition is its intrinsic capacity to resist falsification by relevant variables, while the posterior inductive probability is the falsification-resisting capacity of its implication by certain other factors. So it can easily happen that $p_I[Sa] = 0$ while $p_I[Sa, Ra] > 0$. This is an important property of inductive (Baconian) probability, as distinct from mathematical (Pascalian) probability, in those legal systems or forensic situations in which it is desired to assign a zero prior probability to the guilt of an accused person – the presumption of innocence prior to evidential proof of guilt. The point is that Baconian probability graduates proof on a scale that runs from provability to non-provability, while Pascalian probability graduates it on a scale that runs from provability to disprovability.

VIII

The balance-sheet for neo-Baconian inductive logic therefore looks something like this. On the credit side it sticks closely to the characteristic methods of experimental reasoning in natural science and is responsive to their empirical flexibility. It is quite unrestricted in its application to scientific hypotheses, however rich or complex these may be. It has no difficulty in handling Kyburg's paradox as well as Hempel's and Good-

man's. And it can satisfy Levi's requirement that we should be able to see why people want hypotheses about the world which score well by its criteria. Where certainties are unobtainable, such hypotheses tell us of laws whose operations are relatively resistant to interference. They strengthen our power to plan our lives successfully. But on the debit side ability to compare support-grades in different fields is severely restricted, and it can only rank support-grades, not measure them. Also it cannot be interpreted in terms of familiar Pascalian principles, and seems to complicate our intellectual life by introducing non-Pascalian principles into an area of reasoning where Pascalian principles have often seemed to hold a monopoly of power.

Nevertheless, the first of these three disadvantages is not a very serious one. Neo-Baconian inductive logic is not more modest in its ambitions with regard to inter-field comparisons than are natural scientists themselves.

Also, on the second point it seems in any case preferable to have a ranking-function that is not restricted in application to assessing first-order generalisations about monadic attributes, rather than a measure-function that is so restricted. Moreover it is doubtful whether any measure is possible for support-grades that are evaluated by the toughness of the experimental tests on which they are grounded. This is because the conditions requisite for additivity do not seem to obtain here. The nature of causality is such that the toughness of a test that manipulates two relevant variables does not depend only on the several toughnesses of tests manipulating each variable separately but also on the various kinds of further effects that the variants of the two variables may combine to produce.

The third point, however, raises some deep issues about the nature of probability and about the relationships between Pascalian and non-Pascalian systems. It is certainly clear that if one thinks of probability, from the semantical point of view, as a generalisation on the notion of provability, then, for the following reasons, at least one important kind of non-Pascalian system must be accepted. Just as proof rules may be pigeon-holed in a classificatory matrix in accordance with whether they are singular or general, extensional or non-extensional, and necessary or contingent, so too the familiar forms of Pascalian probability, as relative frequency, logical relation, propensity, degree of belief, etc. may all be seen as generalisations on such different kinds of provability.[25] For example, a relative-frequency conception of probability may be seen as being analogous to provability in a deductive system of which the rules are general, extensional and contingent. But all these forms of generalisation on the concept of provability – even if only comparative and not

measure-theoretic – are based on an assumption of completeness about the proof system, in the sense that h is provable if and only if $-h$ is not provable. Hence arises the familiar Pascalian principle of complementationality: $p[x, y] = 1 - p[-x, y]$. Yet very many proof-systems are *in*complete, in this sense, so that in an analogous system of probability it would be possible to have $p[x, y] = 0 = p[-x, y]$. The generalised scale would then run down from proof to no-proof, rather than from proof to disproof; and such a probability-function would grade how complete the evidence is when on balance it stands in favour of a particular conclusion, rather than how large a proportion of the presumptively complete evidence stands in favour of that conclusion.

So some fairly abstract considerations about the nature of probability expose the existence of a conceptual niche which inductive probability-functions are admirably suited to occupy, since, as we have seen, they grade the weight of evidence. Correspondingly probability-functions that are regulated by Baconian modal logic are just as legitimately entitled 'probability-functions' as the functions regulated by Pascalian mathematics. Indeed, the seeds of both theories were sown within a short while of one another in the seventeenth century, even though Pascalian theory came earlier to maturity. If Pascalian theories have hitherto appeared to monopolise the systematic analysis of probability-judgements, we must put this down to an accident of human intellectual history rather than to a necessary principle of rationality.

The underlying connections, and underlying differences, between the Pascalian inductive logic of Carnap and Hintikka and the neo-Baconian inductive logic that is generated by the method of relevant variables may also be illuminated by reference to the appropriate models that can be constructed in terms of alternative possible worlds, at least so far as the possible-world metaphor is intelligible.

The model suggested by Carnap and Hintikka's inductive logic is straight forward enough, since each state-description may obviously be taken as true for just one possible world. But how is such a model to be constructed for Baconian induction?

A logically possible world, W_1, should be said to be subject to the uniformities of another, W_2, in regard to generalisations of a particular category, if and only if (i) every such generalisation that is true and instantiated in W_2 is also true, whether vacuously or by instantiation, in W_1, and (ii) every such generalisation that is vacuously true in W_2 is also vacuously true in W_1. A logically possible world should be termed 'physically possible' if and only if it is subject to the uniformities of the actual world. A physically possible world would then be termed a t_1 world if and only if every variant of the first relevant variable, and no

variant of any other relevant variable, is instantiated in it. A physically possible world W would be termed a t_2 world if and only if every t_1 world is subject to the uniformities of W and every admissible combination of a variant of the first relevant variable with zero or more variants of the second is instantiated in W, but no variant of any other relevant variable. And so on for t_3, t_4, t_{n-1} worlds. The actual world, or indeed any t_n world, would be assumed to contain a plenitude of relevant events, whereby every physically possible combination of a variant of the first relevant variable with zero or more variants of other relevant variables is instantiated in it. So, for all i and j, each t_i world is subject to the uniformities of each t_j world, where $j \geq i$, as well as to the uniformities of the actual world.

A generalisation may then be said to have at least ith grade inductive reliability if it is true in all t_i worlds, and the characteristic principles.of Baconian induction emerge. For example, since a t_i world is subject to the uniformities of a t_j world where $j > i$, it follows that if one generalisation holds good in all t_i worlds and another in all t_j worlds, their conjunction must hold good in all t_i worlds. Similarly a high grade of inductive reliability is compatible with the existence of anomalies, since a generalisation may hold good in all t_i worlds, for every $i < n$, even though it is falsified in our actual world: the relation of being 'subject to the uniformities of' is not symmetrical. Indeed an experimental test of ith grade throughness can be thought of as a simulated minimal t_i world; and the idealised domain to which some scientific generalisations apply – e.g. the domain of bodies moving in a vacuum – can also be thought of as a t_i world for some appropriate i. Finally, the relation of being 'subject to the uniformities of' constitutes a relation of inductive accessibility or knowability that holds between some possible worlds in the model and not others. Since this relation is transitive and reflexive but not symmetrical, the appropriate formalisation could be expected (in the light of Kripke's work[26] on modal logic) to relate to C.I. Lewis's system S4, as indeed it does. Hence the prior inductive probability (which equals the inductive reliability) of a generalisation can be viewed as a ranking of inductive range, by contrast with Carnap and Hintikka's conception of a proposition's prior mathematical probability as a measure of its logical range. The inductive range of a generalisation is ranked by reference to the fullest kind of physically possible world – i.e. the highest grade of relevantly eventful world – in which the generalisation always holds good, while the logical range of a proposition is measured by reference to the sum of the values severally assigned to the various logically possible worlds in which the proposition holds good. Neo-Baconian inductive logic looks at physically determined qualities of possible worlds:

Carnap's and Hintikka's systems look at logico-linguistically determined quantities of them. That is ultimately how they differ, and that is ultimately why neo-Baconian inductive logic provides a better representation of experimental reasoning in natural science.

Notes

1. Letter to Couring, March 19, 1678, in *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, ed. by C.1. Gerhardt, Vol. 1,1875, p. 195f.

2. D. Hume, *A Treatise of Human Nature*, 1739, Bk I, Pt. 111, Sec. XI.

3. E.g. P. S. de Laplace, *A Philosophical Essay on Probabilities*, tr. F. W. Truscott and F. L. Emory,1951, Pt. 11, Ch. ix.

4. E.g. James Bernoulli, letter to Leibniz of April 20, 1704, in *Leibnizens mathematische Schriften*, ed. by C.I. Gerhardt, Vol. 111,1855, p. 87f.

5. *The Principles of Science: A Treatise on Logic and Scientific Method*, 1874, Vol. l, p. 276ff.

6. *A Treatise on Probability*, 1921.

7. *Le problème logique de l'induction*, 1924, Eng. transl. in *Foundations of Geometry and Induction*, 1930.

8. 'Studies in the Logic of Confirmation', *Mind* liv,1945, p. lff and p. 97f.

9. E.g. I. Scheffler, *The Anatomy of Inquiry*, 1963, p. 289.

10. E.g. J. L. Mackie, 'The Paradoxes of Confirmation' *British Journal for Philosophy of Science* 13 (1963), 265-77.

11. In lectures subsequently published as Chapters II-IV of *Fact, Fiction and Forecast*, 1954. This may be regarded as a modern form of Leibniz's curve-fitting paradox: however many points are given, an infinity of different curves may still be drawn through them. Cf. Leibniz's letter to James Bernoulli of December 3, 1703, in *Leibnizens mathematische Schriten*, ed. by C.1. Gerhardt, Vol. 111,1855, p. 83f.

12. Cf. also the rather more *ad hominem* argument that Goodman's solution is intrinsically incoherent: A. Zabludowski, 'Concerning a fiction about how facts are forecast', *Journal of Philosophy* lxxi (1974), p. 97ff.

13. *Probability and the Logic of Rational Belief*, 1961, pp. 196-9.

14. Cf. in particular his *Logical Foundations of Probability*, 1950. Carnap (ibid. p. 83) attributes the seminal idea of his theory to L. Wittgenstein, *Tractatus Logico-Philosophicus*, 1922. But I. Hacking, *The Emergence of Probability*, 1975, p. 134ff., traces it to Leibniz.

15. In R. Carnap and R. C. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, Vol. 1,1971, p. 227ff.

16. *The Logic of Scientific Discovery*, 1959, For Popper's 'degree of corroboration' as a measure of acceptability cf. pp. 388, 392, 394, etc.

17. 'Towards a Theory of Inductive Generalisation', in *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*, pp. 274-288.

18. E.g. I. Niiniluoto and R. Tuomela, *Theoretical Concepts and Hypothetico-Inductive Inference*, 1973.

19. J. Hintikka and R. Hilpinen, 'Knowledge, Acceptance and Inductive Logic', in *Aspects of Inductive Logic*, ed. by J. Hintikka and P. Suppes,1966, pp. 1-20.

20. *Gambling with Truth*, 1967.

21. Ibid. p. 25.

22. These appeals to intuition are referenced and criticised in L. Jonathan Cohen, 'How Empirical is Contemporary Logical Empiricism?', *Philosophia* 5 (1975), pp. 299-317.

23. *The Structure of Scientific Inference*, 1974. In a more extensive coverage it would also be necessary to discuss those writers (e.g. H. Reichenbach, *The Theory of Probability*, 1949, p. 429ff.) who take induction about scientific generalisations to be just the limiting-case of some mode of statistical reasoning about non-universal correlations. Their general fault is to attach greater value to enumerative induction than is compatible with good scientific practice.

24. This theory was first presented in detail in L. Jonathan Cohen, *The Implications of Induction*, 1970, and was further extended (with some minor revisions) in *The Probable and the Provable*, 1977.

25. Cf. L. Jonathan Cohen, *The Probable and the Provable*, 1977, Secs. 1-9.

26. S. Kripke, 'Semantical analysis of modal logic I, normal propositional calculi', *Zeitschrift fur mathematische Logic and Grundlagen der Mathematik* 9 (1963), 67-96.

16

SOME HISTORICAL REMARKS ON THE BACONIAN CONCEPTION OF PROBABILITY*

According to Ian Hacking, Francis Bacon had "no concern with probability" and "does not aim at inference under uncertainty."[2] I believe this to be an important mistake, though such mistakes are rare in Hacking's fascinating book. In fact Bacon, and later writers influenced by him, were very much concerned with probabilities, though not with probabilities structured in accordance with the mathematical calculus of chance. I shall call the latter "Pascalian probabilities," in tribute to one of the great mathematical pioneers in this area; and my object will be to demonstrate not only Bacon's own concern with a non-Pascalian probability, but also the existence of a long line of philosophical or methodological reflections about such a probability, stretching at least from the seventeenth into the nineteenth century.

1. Bacon's Own Views

During the past hundred years Bacon's writings have been seriously misinterpreted with surprising frequency. In particular he is often supposed to have put forward a method of investigating nature whereby scientists might proceed at one leap, as it were, from singular observations to knowledge of general truth. "Absolute certainty," wrote Ellis,

*Earlier versions of this paper were read at the XVth International Congress of the History of Science, in Edinburgh (August 1977) ; at the Royal Institution (History of Science Discussion Group) in London (February 1978); and at the IInd International Conference on History and Philosophy of Science in Pisa (September 1978). I am grateful to other participants at these meetings for helpful comments and criticisms.

"is ... one of the distinguishing characters of the Baconian induction."[3] Indeed Bacon's own words sometimes tend to suggest this interpretation, as in the passage quoted by Hacking[4] where Bacon says that his "course and method ... is to extract causes and axioms from works and experiments." Hacking concludes here that "there is little room in this conceptual scheme for a working concept of probability." Similarly, according to L. Jardine[5] Bacon believed that as a result of a single induction "it would be immediately evident which natures fall into groups by virtue of the shared fundamental terms in their forms."

She thinks it wrong to interpret Bacon as holding that there are successive levels of forms, or causes, and that after the form of, say, whiteness has been discovered a further induction will explain this form in terms of yet more fundamental qualities. Consequently, on Jardine's view as on Hacking's, there is no room for laws established by Baconian induction to constitute a pyramid of increasing probability. And again Bacon's words sometimes fit this interpretation. He certainly did believe that there were relatively few forms of the first class, as he called them, and he said that these were responsible, in their various combinations, for all the forms and essences of ordinary substances.[6] But it is quite clear from other passages that Bacon was just summarizing here and that in fact he envisaged a hierarchy of forms to exist in Nature, with a consequential opportunity for investigators to increase the reliability of their findings as they arrive at higher and higher laws.

For example, Bacon enunciated the precept that, in order to find the form for a particular nature, "another nature should be discovered which is convertible with the given nature and yet is a limitation of a more intelligible nature, like a true natural kind." [7] And we have to suppose that such a more fundamental nature may itself have a form that investigators can seek to discover, since Bacon's own examples imply this. For instance, he says[8] that the nature of which heat is a limitation is motion: the form of heat, or definition of its essence, is expansive motion that is restrained and acting in its strife upon the smaller particles of bodies. But elsewhere[9] he treats motion itself as a nature that may be investigated in the same way as heat is investigated. Again, Bacon insists[10] that in Nature nothing really exists besides individual bodies performing individual acts according to law, and that it is this law, with its clauses, which he understands when he speaks of "forms." The form of heat or the form of light is nothing but "the law of heat" or "the law of light," knowledge of which enables a man to superinduce heat or light, as the case may be, under suitable conditions.[11] So a hierarchy of Baconian forms is a hierarchy of causal laws. Such laws are desribed in what Bacon calls axioms;[12] and he speaks of a gradual ascent to

more and more general axioms.[13] The structure of scientific knowledge is thus that of a pyramid.[14] Indeed, to any contemporary of Bacon's the Platonic and Aristotelian echoes of his terminology could hardly have failed to suggest a hierarchy of true forms or natural kinds (and thus of laws) rather than a single level. And, if Bacon had intended to disallow this suggestion, one would scarcely have expected him to reinforce it with his talk about a pyramid and about gradual ascent.

Moreover even at any one level of the hierarchy the inductive method by which an axiom is established is essentially a gradualist procedure. The tables of presence and absence that form the basis for eliminative induction are not easily completed. Bacon's method requires the successive exclusion, as J. S. Mill put it,[15] of the various circumstances which are found to accompany a phenomenon in a given instance, in order to ascertain what are those among them which can be absent consistently with the existence of the phenomenon. Indeed Bacon laid rather more stress on this point that Mill did, because Mill failed to acknowledge that evidence gained by varying some kinds of circumstance may have greater significance than evidence gained by varying others. We do not learn much more about the causal powers of arsenic from the fact that women as well as men die after drinking it. But Bacon clearly thought that some types of observations, which he called "prerogative instances," had much greater inductive value than what he called "common instances." And he seems to have arranged his list of twenty-seven such types of observation in the order of importance that he attached to them. For example, the "solitary instances" that he placed first are such as to "make the way short, and accelerate and strengthen the process of exclusion."[16] It is then obviously possible to grade the reliability of a hitherto unfalsified axiom, in Bacon's methodology, by the extent to which it has been tested down the list of prerogative instances. Bacon does not in fact say this. But what he does say, at the very beginning of the second paragraph of his preface to the *Novum Organum*,[17] is that his principle of method, "though hard to practise, is easy to explain: it involves setting up degrees of certainty." ("Nostra autem ratio, ut opere ardua, ita dictu facilis est. Ea enim est, ut certitudinis gradus constituamus.") The only way to interpret this programmatic announcement is in terms of what he says later about the order of prerogative instances and the ladder of axioms.

Indeed, though again Bacon does not make the point explicitly, these two stages in the inductive process have essentially the same logical structure. Both levels of reasoning–both initial experimental induction and also the merging of relatively specific uniformities into more comprehensive ones–proceed by assigning greater reliability or certainty to

the hypothesis that holds good over a bigger range of appropriately selected variations. On the one hand, the connection between a "nature" (or phenomenon) and its "form" (or cause) is strengthened by there being instances both of their co-presence and of their co-absence over a cumula tively wider and wider variety of relevant circumstances: on the other hand, the reliability of an axiom (or law) is strengthened by its prevalence in a cumulatively wider and wider variety of particular manifestations.

2. The Views of Some Baconians.

What seems to have led Ellis, and many others after him, astray is Bacon's insistence that his method was capable of producing conclusively certain results. Bacon's modern critics have tended to infer from this, quite wrongly, that Bacon thought such results to be immediately obtainable by his method in any area of enquiry. But in fact Bacon stated explicitly, as we have seen, that on his view the ascent to certainty was a difficult one and passed through many distinguishable stages. Of course, Bacon was wrong to suppose that his method could even in the end produce conclusively certain results. His mistake here sprang from a failure to recognize that in eliminative induction every prior assumption about the variety of hypotheses that are open to elimination is itself empirically corrigible. We can never be conclusively sure that our list of forms, natures, prerogative instances, or whatever, is complete. It may turn out later that a hidden variable was operating in our experiments. So that any probabilities or certainties that we may be inclined to affirm are relative to prior assumptions about our space of alternative hypotheses. But Bacon's mistaken belief in the possibility of eventual certainty should not be allowed to obscure his correct belief in the possibility of grading stages in the elimination of legitimate doubt.

There is no doubt at all that many of Bacon's seventeenth-century admirers were gradualists; and the fact that they never felt their gradualism to be in any way inconsistent with Bacon's philosophy is further confirmation that the interpretation proposed by Ellis, Hacking, Jardine, etc. is mistaken. The precise terminology does not matter here. Whether we are thinking in Pascalian or in non-Pascalian terms, we can speak either of degrees of probability, with an upper limit describable as certainty, or of degrees of certainty with an upper limit describable as certainty simpliciter. Bacon seems to have preferred the latter locution, but there is no significance in this preference, and his intellectual posterity often chose the other idiom. What is important is not the choice of

vocabulary, but the underlying idea that the reliability of certain rules of inference (those given by causal laws) is a matter of degree.

For example, the great experimentalist Robert Hooke thought that Bacon was the only man who had developed a correct methodology of scientific enquiry,[18] and he wished to build on the foundations that Bacon had laid. But Hooke spoke of the results obtained hitherto by this method as "probabilities,"[19] though he himself, like Bacon, was convinced that demonstrations were possible in physical as well as in mathematical enquiry.

Robert Boyle was another admirer of Bacon. He too held that those who desire to command nature, not only to know her, need to make experiments and observations in appropriately varied circumstances.[20] He too believed in a hierarchy of explanations, claiming that[21]

> To render a reason of an effect or phaenomenon, is to deduce it from something else in nature more known than itself. ...The nearer the discovered causes are to those, that are highest in the scale or series of causes, the more is the intellect both gratified and instructed.

And when he discusses a medical remedy that will work in some circumstances but not in others he speaks of its "uncertainty."[22]

Yet another admirer of Bacon, Joseph Glanvill, who thought[23] that Solomon's House in Bacon's *New Atlantis* was prophetic of the Royal Society, emphasized[24] that natural scientists were wise to claim probability, not certainty, for their hypotheses. And this probability was subject to gradation. A man should proportion the degree of his assent, said Glanvill,[25] to the weight of the evidence. The more thoroughly the theory had been examined and the more clearly it was seen to work, the higher its probability.

3. The Recognition of Different Types of Probability.

Bacon and his disciples were thus much concerned with the problem how the truth of a probability-judgment was determined or with the circumstances under which one inference or conclusion was to be judged more probable than another. But they gave very little consideration to the logical or mathematical relations between such probability-judgments.

It is clearly *implicit* in Glanvill's remarks (cited above) that the relevant scale of probability runs upwards from no evidence, or non-proof, to proof, rather than, as in any Pascalian scale, from disproof to proof. Indeed the former is the natural way to think of it, if we focus our minds on the process by which alternative hypotheses are gradually eliminated and the range of conclusions with which the evidence is compatible is gradually narrowed down. And we can find similar non-Pascalian con-

ceptions of probability being used in later centuries, though often without any Baconian overtones. For example, Joseph Butler remarked on the first page of his introduction to *The Analogy of Religion* (1736)

> Probable evidence is essentially distinguished from demonstrative by this, that it admits of degrees; and of all variety of them, from the highest moral certainty, to the very lowest presumption.

And a couple of pages later he added

> In questions of difficulty, or such as are thought so, where more satisfactory evidence cannot be had, or is not seen, if the result of examination be, that there appears upon the whole any the lowest presumption on one side, and none on the other, or a greater presumption on one side, though in the lowest degree greater; this determines the question.

In the above passage Butler clearly endorses a non-Pascalian conception of probability, since he speaks of a possible state of affairs in which prob(B/A)–the probability of B, given A–may be greater than zero though substantially less than unity while prob(not-B/A)–the probability of not-B, given A–is equal to zero. Such a possibility is incompatible with the familiar complementational principle for negation—prob -B/A = 1-prob B/A—that the Pascalian calculus generates. No doubt some people will be tempted to object that Butler just made a mistake here. But historians need to be very careful about making normative judgments of this kind. The fact is, as we have already seen, that if probability be regarded as a gradation of provability, which is just how Butler does regard it, there is undeniably room for two different kinds of scale, depending on whether the lower extreme is taken to be the level of disprovability or of non-provability. Butler, with his talk of comparative presumptions, has adopted the latter kind of scale, and quite consistently he rejects a complementational principle for negation.

This conception of probability is particularly appropriate for the evaluation of opposing proofs in legal proceedings. James Glassford, an early nineteenth-century Scots advocate, considered[26] that

> the scale of Evidence, according to the Degrees of probability which it furnishes, from the lowest upward, may, upon a comparison or medium of ordinary cases, be stated as follows, viz. Evidence of General Presumptions—Evidence of a sole Witness, or of a Writing proved by a single testimony—Circumstantial Evidence—Evidence of a sole Witness, or of Writing so proved, and Circumstantial evidence combined—Evidence of two or more Witnesses, or of Writing so proved, and Circumstantial evidence, combined.

Here too the scale of probability obviously ran upwards from non-proof, not from disproof.

We must apparently look to David Hume, however, for the first explicit recognition by one of Bacon's admirers that there is an important kind of probability which does not fit into the framework afforded by the

calculus of chance. Hume, who subtitled his *Treatise of Human Nature* "An attempt to introduce the experimental method of reasoning into moral subjects," regarded Francis Bacon as being the founder of empirical science in the modern world,[27] and comparable thus to Thales in the ancient one. Admittedly Hume's epistemological skepticism mingles awkwardly with the scientific ambitions manifested in his psychological theory-construction. The associationist theory is itself open to the same skeptical challenge as any other causal hypothesis, though it is needed by Hume in order to explain the existence of beliefs that cannot be explained by there being rational grounds for them. And Hume's skeptical arguments are aimed against the possibility of obtaining knowledge by enumerative induction, though his own "Rules by which to judge of causes and effects."[28] are typically Baconian in character. Hume certainly had a potentially incoherent ambition to hunt with the skeptical hounds at the same time as he ran with the scientific hare. But this should not blind us to the perceptiveness of his distinction (which Hacking does not mention) between probabilities arising from analogy and probabilities arising from chance or cause.

Probabilities arising from chance Hume evaluated by the ratio of favorable outcomes among a class of outcomes that are all indifferently possible.[29] Any one side of a six-sided die has a 1/6 probability of falling uppermost. And probabilities arising from cause he evaluated by what Carnap called the straight rule. To use his own example: if I have found by long observation that of twenty ships which go to sea only nineteen return, then 19/20 measures the strength of my belief that a ship which goes to sea in the future will return.[30] Hume clearly thought that the general nature of human reasoning about probabilities arising from chance and probabilities arising from cause was the same:

> Every past experiment may be consider'd as a kind of chance; it being uncertain to us, whether the object will exist conformable to one experiment or another: And for this reason every thing that has been said on the one subject is applicable to both.[31]

But probabilities arising from analogy, according to Hume, are substantially different. John Locke had already stressed the importance of analogy in connection with probability. He regarded analogy as "the great rule of probability" in "things which sense cannot discover."[32] But Hume had a deeper understanding of the matter than Locke. Hume thought that any kind of reasoning from causes and effects was founded on two factors: the constant union of type A and type B situations in all past experience, and a present situation's resemblance to one of them. If you diminish either the constancy of the union or the extent of the resemblance, you thereby weaken the impact on an observer's beliefs and reduce a feeling of certainty to one of mere probability. In the case of probabilities arising from chance or from cause, it is the constancy of the union that is diminished. But probabilities arising from analogy are

different because in their case it is the extent of the resemblance that is diminished:[33]

> Without some degree of resemblance, as well as union, 'tis impossible there can be any reasoning: but as this resemblance admits of many different degrees, the reasoning becomes proportionally more or less firm and certain. An experiment loses of its force, when transfer'd to instances, which are not exactly resembling; tho' 'tis evident it may still retain as much as may be the foundation of probability, as long as there is any resemblance remaining.

Hume never suggests, however, that probabilities arising from analogy may be measured, like those arising from chance or cause, by a numerical ratio. And in fact it is clear that they must be assigned a non Pascalian structure, because the negation principle for such probabilities must be a non-complementational one. It cannot have the form: p(B/A) =1-p (not-B/A). So far as in relevant respects the circumstances of the present situation partially resemble, and partially differ from, those of a previous situation which was immediately followed by an event of type B, we have evidence for only a probability, perhaps only a very slight probability, that an event of type B will follow the present situation. But we certainly have not got evidence thereby for a positive probability. perhaps quite a high positive probability, that an event of type B will not happen. To have evidence for *that* conclusion we should need to know not merely that the circumstances of the present situation partially resemble those of a previous situation which was followed by an event of type B, but also that the circumstances of the present situation partially resemble those of a previous situation which was not followed by an event of type B. And it is easy enough to know the former, in a particular case, without also knowing the latter. In other words analogical probability seems to belong naturally to a scale that spans the interval between provability and non-provability–between adequate evidence and no evidence–rather than to one that has a complementational principle for negation because it runs from provability to disprovability.

Nor should it be surprising that in this respect analogical probability shares a key non-Pascalian feature with Baconian induction. The strength of a piece of analogical reasoning about nature lies wholly in the infra-structure of eliminative induction that it presupposes. The intimate links between analogical and Baconian reasoning were well brought out by J. S. Mill: [34]

> In the strictest induction, equally with the faintest analogy, we conclude because A resembles B in one or more properties, that it does so in a certain other property. The difference is, that in the case of a complete induction it has been previously shown, by due comparison of instances, that there is an invariable conjunction between the former property or properties and the latter property; but in what is called analogical reasoning, no such conjunction has been made out. There have been no

> opportunities of putting in practice the Method of Difference, or even the Method of Agreement; but we conclude (and that is all which the argument of analogy amounts to) that a fact m, known to be true of A is more likely to be true of B if B agrees with A in some of its properties. ... The supposition must be that m is an effect really dependent on some property of A, but we know not on which. ... There can be no doubt that every such resemblance which can be pointed out between B and A affords some degree of probability, beyond what would otherwise exist, in favour of the conclusion drawn from it.

But Mill went on to claim that, if after much observation of B, we find that it agrees with A in nine out of ten of its known properties, we may conclude with a probability of nine to one, that it will possess any given derivative property of A.[35] Rather naively he assumed not only that all the properties involved would be of equal significance but also that they would be additive, so that he would be entitled to measure probability here by an arithmetical ratio. In fact both assumptions are highly precarious. The various factors relevant to testing a particular causal hypothesis may differ very much from one another in importance, as Bacon acknowledged in his discussion of prerogative instances. And the Baconian probability of a generalization, since it is determined by the generalization's capacity to resist falsification under possible combinations of relevant causal factors, is inherently non-additive. Two causal factors may have an explosive (or mutually neutralizing) effect in combination which neither even remotely approaches in isolation.[36] Hume was correspondingly right not to suggest using arithmetical ratios to evaluate the probabilities of analogical reasoning.

J. F. W. Herschel also, who set an engraving of Bacon's bust upon the title page of his *A Preliminary Discourse on the Study of Natural Philosophy*,[37] had an intuitive picture of the situation. He certainly speaks of probability in two rather different ways. At one point[38] he identifies probability with what is studied by

> a refined and curious branch of mathematical enquiry, called the doctrine of probabilities, the object of which ... is to reduce our estimation of the probability of any conclusion to calculation, so as to be able to give more than a mere guess at the degree of reliance which ought to be placed on it.

But the problem with which Herschel is exercised at this point is just that of maximizing the accuracy with which a physical quantity may be measured. Where single observations give different values, he says, the average of all observations (under certain restrictions) will be the exact value; and where certainty is not to be had–because we cannot exhaust the totality of possible observations–the most probable value is the one to be adopted. Appropriately enough Herschel adopts here a Pascalian

conception of probability. And it is only in connection with this problem of measurement, he adds, that he has anything to say about probability.

In fact, however, Herschel had aready said something else about probability some seventy pages earlier, in quite a different context. He was there discussing the typically Baconian problem–a problem not of measurement, but of the search for causes:[39]

> Whenever ... any phenomenon presents itself for explanation, we naturally seek in the first instance, to refer it to some one or other of those real causes which experience has shown to exist, and to be efficacious in producing similar phenomena. In this attempt our probability of success will, of course, mainly depend, 1st, On the number and variety of causes experience has placed at our disposal; 2dly, On our habit of applying them to the explanation of natural phenomena; and, 3dly, on the number of analogous phenomena which have either been explained or which admit of explanation by some one or other of those causes, and the closeness of their analogy with that in question.

Now Herschel was not only deeply committed to eliminative induction as the correct method of scientific enquiry. He also saw clearly, like Bacon, that the success which such an investigation of a particular problem had achieved at any one time was essentially a matter of degree:[40]

> In nature, it is comparatively rare to find instances pointedly differing in one circumstance and agreeing in every other; but when we call experiment to our aid, it is easy to produce them; and this is, in fact, the grand application of experiments of enquiry in physical researches. They become more valuable, and their results clearer, in proportion as they possess this quality (of agreeing exactly in all their circumstances but one), since the question put to nature becomes thereby more pointed, and its answer more decisive.

But the decisiveness of an experiment, in this sense, increases with the number of different kinds of relevant circumstances that are being varied in the experiment. Equally we should expect the replicability of an experiment to depend on this consideration. The more an experimenter knows about the relevant factors which are operative in a detected result, the greater the probability that his attempts to replicate the result will succeed or that his hypothesis about what caused the result is correct. Herschel's text therefore suggests that he implicitly recognized a difference between the type of probability that graduates eliminative induction and the type of probability that is regulated by the mathematical calculus of chance.

J. S. Mill's failure to recognize this difference is obviously connected with his failure to admit any graduation at all for eliminative induction. Mill did not acknowledge, as Herschel acknowledged, that the decisiveness of an experiment is normally a matter of degree. So he gave himself no opportunity to study the structure of this graduation. Instead his five canons of experimental enquiry were formulated as statements of

methods by which the cause or effect of a given phenomenon could at one go be *discovered* or *proved*. Indeed it may well be that Mill's influential formulations tended to foster a conception of inductivist doctrine which made it easier for Ellis and others to misinterpret Bacon. Anyone who thought that Mill was developing Bacon's ideas would be liable to make the mistake of imputing absolute certainty to Baconian induction.

4. The Relevance of Modal Logic

In sum, though the story is a tangled one, it is undoubtedly wrong to say that Bacon and the Baconian tradition had no concern with probability. Equally it would be wrong to say of the classical Baconians (with the single exception of J. S. Mill) that they thought of eliminative or analogical probability as having anything to do with the mathematical calculus of chance. But they did not put forward any positive account of the constraints that one such judgment of probability places on another. The Baconian tradition was much concerned with the truth-conditions of such judgments, but apparently never attempted to formulate the logic of their interrelations. No explicit counterparts were forthcoming for the familiar Pascalian principles. Specifically, the multiplicative law for conjunction, complementational law for negation, etc. were not thought to apply to Baconian probability. But it was not said what principles did govern conjunction, negation, etc. in that case.

Why was this? The only plausible answer to the question is that the right kinds of logical or mathematical tools were not available until the present century. Pascal, Fermat, Leibniz, Huyghens, Bernoulli, etc. could build on the foundations of combinatory arithmetic that were already available. But Baconian probability had a very different structure.

It is easy enough, with hindsight, to discern the outlines of this structure. For example, since a conjunction of Baconian axioms survives every variation of circumstance that both its conjuncts survive, the certainty of a conjunction A & B must be equal to that of A, if the certainty of A is less than or equal to that of B. Again, a Baconian axiom connecting a form F with the presence of a nature N can only acquire certainty through its surviving variations of circumstance in which a hypothesis connecting F with the absence of N acquires no certainty at all. So any positive Baconian probability for N, given F, implies a zero Baconian probability for not-N, given F. But it turns out[41] that a systematic formulation of such principles requires a generalization of C. I. Lewis's modal logic S4, which demonstrably resists mapping onto the calculus of chance. Indeed we can think of a controlled experiment, in which a situation is supposed to be insulated from every causal factor except

those duly controlled, as a rather jejune possible world in which only the relevant kinds of circumstance occur. Generalizations of the appropriate kind then attain higher and higher levels of inductive probability as they hold good in richer and richer worlds. But, as they approach thus closer and closer to holding good in all physically possible worlds, they approach closer and closer to attaining natural necessity. So it is scarcely surprising that the logic of controlled experiment treats inductive probability as a generalization on the notion of natural necessity. And the backwardness of modal logic until the present century therefore constituted an effective bar to any earlier formulation of the constraints that Baconian probability judgments impose on one another.

No doubt it will be said that the term "probability" has long been preempted for the Pascalian tradition, and that one should therefore not use it in connection with other ways of grading inferences or propositions. But this confuses what is appropriate for certain kinds of mathematical enquiry with what is appropriate for any historical or philosophical enquiry. Once the common mathematical structure of certain concepts has been grasped (chance, relative frequency, coherent betting quotients, etc.) it is useful to give them a common name in virtue of that shared structure. The structure can be axiomatized, and the mathematical properties of appropriate axiom-systems can be investigated, under a convenient label. But the interest of such investigations should not be allowed to blind us to the existence of other important ways in which the term "probability" functions or has functioned. For example, if we are engaged on the philosophical investigation of contemporary reasoning, we shall find a strong case for putting a non-Pascalian interpretation upon the concept of probability that is used in stating the standards of proof in Anglo-American law.[42] And if we are engaged on historical research, we can find-as I have shown above–a long tradition of theorizing about probability within an implicitly non-Pascalian framework. Indeed this systematic ambiguity of the term "probability" needs to be clearly recognized in order to help account for the kind of confusion into which J. S. Mill (and many after him) fell. Moreover, from a semantical point of view both Pascalian and non-Pascalian uses of the term "probability" can be explicated within a common framework. If probability is a generalization on the notion of provability or inferability, as so many philosophers have held,[43] then Pascalian probability, with its complementational principle for negation, generalizes on the notion of inferability within a complete deductive system, where a system is said to be complete if a proposition A is provable within it if and only if not-A is not provable. But if we generalize on the concept of inferability within a system that is not complete, in this sense, then we obtain a con-

cept of probability that has to have a non-complementational principle for negation. So an inferentialist semantics for the term "probability" comes naturally to generate both Pascalian and non-Pascalian readings for that term. To suppose that the Pascalian reading is the only intellectually respectable one is like supposing that the only intellectually respectable concept of space is the Euclidean one.[44]

Notes

1. Earlier versions of this paper were read at the XVth International Congress of the History of Science, in Edinburgh (August 1977) ; at the Royal Institution (History of Science Discussion Group) in London (February 1978); and at the IInd International Conference on History and Philosophy of Science in Pisa (September 1978). I am grateful to other participants at these meetings for helpful comments and criticisms.

2. I. Hacking, *The Emergence of Probability* (Cambridge, 1975), 76.

3. R. L. Ellis, "General Preface to Bacon's Philosophical Works," in *The Works of Francis Bacon*, ed. J. Spedding, R. Ellis, and D. N. Heath (London, 1879), 1, 23.

4. *Op. cit.*, 27, from *Novum Organum*, Bk.I, sec.cxvii, in *Works*, I, 212.

5. *Francis Bacon, Discovery and the Art of Discourse* (Cambridge, 1974), 130.

6. *De Augmentis Sciendarum* Bk.III, sec.iv, in *Works*, I, 566-68.

7. *Novum Organum*, Bk.II, sec.iv, in *Works*, I, 230.

8. *Ibid.*, Bk.II, sec.xxxv, in *Works*, 1, 261ff.

9. *Ibid.*, Bk.II, sec.xxxv, in *Works*, I, 291.

10. *Ibid.*, Bk.II, sec.ii, in *Works*, I, 228.

11. *Ibid.*, Bk.II, sec.xvii, in *Works*, I, 257-58.

12. *Ibid.*, Bk.II, sec.v, in Works, I, 230-31.

13. *Instauratio Magna, Distributio Operis*, in *Works*, I, 136-37, and *Novum Organum*, Bk.II, sec.civ, in *Works*, I, 205.

14. *De Augmentis Scientiarum*, Bk.III, sec.iv, in *Works*, I, 568.

15. *A System of Logic, Ratiocinative and Inductive* (London, 1896), Bk.III, ch.viii, sec.3, 256.

16. *Novum Organum*, Bk.II, sec.xxii, in *Works*, I, 268.

17. *Works*, I, 151.

18. "A General Scheme, or Idea of the Present State of Natural Philosophy and How its Defects may be Remedied by a Methodical Proceeding in the making of Experiments and Collecting Observations, whereby to Compile a Natural History, as the Solid Basis for the Superstructure of True Philosophy," in *The Posthumous Works of Robert Hooke*, ed. R. Waller (London, 1705), 6.

19. *Ibid.*, 7.

20. "New Experiments Physico-mechanical touching the Spring of Air, and its Effects; made for the most part in a new Pneumatical Engine," in *Works* (London, 1744), 1, 199.

21. *Ibid.*, 198.

22. "Two Essays Concerning the Unsuccessfulness of Experiments," op. cit., 222. Boyle seems, however, not to have learned from Bacon how important it was to use adequate controls in assessing the efficacy of medicines that he gave to patients. Cf. L. S. King, "Robert Boyle as an Amateur Physician," in C. W. Bodemer and L. S. King, *Medical Investigation in Seventeenth-Century England* (Los Angeles, 1968), 43

23. *Scepsis Scientifica* (London, 1665), introductory address.

24. *Essays on Several Important Subjects in Philosophy and Religion* (London, 1676),44.

25. *Ibid.*, 46. On the distinction in seventeenth-century England between different kinds of certainty (mathematical, sensory, religious, natural-scientific, etc.) cf. H. G. van Leeuwen, *The Problem of Certainty in English Thought 1630-1690* (The Hague, 1970): this distinction does not concern us here.

26. *An Essay on the Principles of Evidence and their Application to Subjects of Judicial Enquiry* (Edinburgh, 1820), 653.

27. *A Treatise of Human Nature* (London, 1739), Introduction: ed. L. A. Selby-Bigge (Oxford, 1888), xx-xxi.

28. *Ibid.*, Bk.I, pt.III, sec.XV, 173-76.

29. *Ibid.*, Bk.I, pt.III, sec.VI, 125ff.

30. *Ibid.*, Bk.I, pt.111, sec.XII, 134f.

31. *Ibid.*, 135.

32. *An Essay Concerning Human Understanding*, 5th ed. (London, 1706), Bk.IV, ch.XVI, sec.12.

33. Hume, *op. cit.*, 142.

34. *Op. cit.*, Bk.III, ch.XX, sec. 1, 367. Mill severely criticized Sir William Hamilton for ignoring these links: cf. J. S. Mill, *An Examination of Sir William Hamilton's Philosophy and of the Principal Philosophical Questions Discussed in his Writings* (London, 1865), 420f.

35. *Op. cit.*, Bk.III, ch.XX, sec.3, 367.

36. The argument for the non-additivity of Baconian probability is developed at greater length in L. Jonathan Cohen, *The Probable and the Provable* (Oxford, 1977), 226ff.

37. London, 1830.

38. *Ibid.*, 217.

39. *Ibid.*, 148.

40. *Ibid.*, 155. Another nineteenth-century Baconian who seems to have used both Pascalian and non-Pascalian concepts of probability, though without clearly distinguishing between them, is F. Oersterlen; cf. his *Medical Logic*, trans. G. Whitley (London, 1855), 287: "Our calculation of *probability*, equally with our hypotheses and conjectures, will gain in *certainty* the clearer our perception of the occurrences and circumstances with which we have to deal becomes" (my italics). An earlier and much more famous German philosopher who used both concepts without distinguishing them was Leibniz. Leibniz's discussions of Pascalian probability are well-known. But his remarks to Conring (in a letter of March 19, 1678) have a decidedly Baconian flavor. He says nothing there about the partitioning of possibilities, but remarks instead that a hypothesis is more probable ("probabilior") in proportion as it is intellectually simpler though more powerful, i.e., capable of explaining more phenomena on fewer assumptions: *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, ed. C. J. Gerhardt (Berlin, 1875), I, 195-96. In this passage Leibniz endorses, as Bacon had often done (e.g., *Novum Organum* Bk.I, sec.ciii, cvi, cxvii, in *Works*, I, 204, 206, 212, respectively), the importance of a hypothesis' generating novel predictions. It is possible to show how this importance is a necessary consequence of adopting Baconian patterns of inductive reasoning: cf. L. Jonathan Cohen, *op. cit.*, 159-60.

41. For the details see L. Jonathan Cohen, *op. cit.*, 188ff. and 229ff., and also *The Implications of Induction* (London, 1970), 207ff. I am attributing to Bacon only the germs of the logic of controlled experiment that I have developed in these two books. Obviously his own formulations leave many critical issues unresolved.

42. L. Jonathan Cohen, *The Probable and the Provable*, 49-120.

43. *Ibid.*, 13-47.

44. It was only after writing this paper that I had an opportunity to read Glenn Shafer, "Non-Additive Probabilities in the Work of Bernoulli and Lambert;" *Archive for History of Exact Sciences* **19** (1978), 309-70. In his interesting paper Shafer has shown that both Bernoulli and Lambert investigated the mathematics of non-complementational probabilities (called by Shafer "non-additive probabilities"). But they did so, I think, within a framework of assumptions about the additivity (i.e., quantitative measurability) of evidential values which were quite rightly not shared by the classical Baconians (except J. S. Mill) in their concern with the evidence of controlled experimentation.

17

TWELVE QUESTIONS ABOUT KEYNES'S CONCEPT OF WEIGHT

Introduction: Keynes's concept of weight

1 *How Does Weight Come to Matter?*

2 *Why Did Keynes Not Appreciate How Weight Matters?*

3 *Is Weight Increased by Each Addition of Relevant Evidence?*

4 *Do Arguments Inherit Weight Via The Entailments of Their Conclusions?*

5 *Can Arguments in Co-ordinate Terms be Compared for Weight?*

6 *Does Weight Have Any Limiting Cases?*

7 *If One Premiss is of Greater Relevance Than Another, Does it Add More Weight?*

8 *Is Weight Determined by Related Probabilities?*

9 *Can One Argument be Compared for Weight With Another if its Terms Neither Entail nor are Co-ordinate with the Other's Terms and Neither Argument is a Limiting-Case?*

10 *Can Weight be Ranked as Well as Compared?*

11 *Is it Worth While Knowing the Weight of an Argument Without Knowing its Probability?*

12 *What is the Connection Between Keynesian Weight and Baconian Legisimilitude?*

INTRODUCTION: KEYNES'S CONCEPT OF WEIGHT

In chapter VI of his [1921] Keynes treats the probability of H on E as a property of the argument from E to H. The probability depends for its value on the balance between the favourableness and unfavourableness of the evidence that E states in relation to H. But he considers that there may be another respect in which some kind of quantitative comparison between arguments is possible. 'This comparison turns,' he says, 'upon a balance, not between the favourable and the unfavourable evidence, but between the *absolute* amounts of relevant knowledge and relevant ignorance. As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but *something* seems to have increased in either case,–we have a more substantial basis upon which to rest our conclusion.' Keynes expresses this by saying that an accession of relevant evidence increases what he calls the 'weight' of an argument. Thus the weight of an argument is independent of the correctness or incorrectness with which such-or-such a probability is assigned to the argument and is not necessarily determined by the probable error of the argument's conclusion (where that conclusion assigns a value to a magnitude). Keynes says that, metaphorically, the weight of the argument from E to H measures the *sum* of the favourable and unfavourable evidence that E states for H, and the probability measures the *difference*. But he does not suggest any method by which weights might be measured and in fact admits that often one cannot even compare the weights of different arguments. He thinks that, 'in deciding on a course of action, it seems plausible to suppose that we ought to take account of the weight as well as the probability of different expectations'. But he finds it difficult to think of any clear example of this, and he does not feel sure that the theory of evidential weight has much practical significance.

In this paper I shall raise twelve questions about weight, in Keynes's sense; and by answering those questions I shall try to show not only why a theory of evidential weight is needed, but also what form it should take if Keynes's seminal intuition is to be preserved. I shall not, however, use Keynes's own symbolism. Keynes used the letters 'h', 'h_1', 'h_2', etc. for the premisses of an argument, not its conclusion–which is confusing to a contemporary reader. So I shall use 'E', 'E_1', 'E_2', etc. for premisses and 'H', 'H_1', 'H_2', etc. for conclusions. Also he abbreviated 'the probability of the argument from h to a' as 'a/h', whereas I shall use instead the more conventional formula '$p(H/E)$' for the probability of the argument

from E to H. But I shall follow Keynes in speaking sometimes of the weight of the argument from, say, E to H, sometimes of the weight of the evidence E for H, and sometimes of the weight of the probability of H on E. This flexibility is harmless so long as one remembers that by speaking of the weight of the probability of H on E Keynes does not intend to treat weight as a property of certain propositions such as of the proposition $p(H/E) = n$. On his view it is possible to know the weight of a probability without knowing its value, or to know its value without knowing its weight.

1. HOW DOES WEIGHT COME TO MATTER?

In calculating the premium that a client should pay for a life insurance policy maturing at age 65 a company that wanted to minimise the risk of a loss on this class of business would ideally determine the probability of the client's death before the age of 65 on the basis of all the relevant facts, *i.e.* of all the facts that affect the probability one way or the other. In practice it may well be uneconomic to enquire too closely into the client's health, ancestry and life-style, and some of the relevant facts (*e.g.* about grandparents' medical histories) may be quite unobtainable. But at least the company needs to know the client's sex, say, and whether he or she is at present ill or has a particularly dangerous job or hobby. Since people who have a reason to fear early death are particularly likely to want life insurance protection in the interests of their dependents, a company risks bankruptcy if it does not take such reasons into account in its calculation of premiums to be charged. Equally it risks losing business to other companies if it does not offer appropriate reductions in premiums to clients who show obvious prospects of exceptional longevity. In other words the company must in each case have an appropriate weight of evidence for the probability of survival to 65 that is accepted as the basis for calculating an economic premium. Since an individual may have many relevant features that constitute items of favourable or unfavourable evidence about his survival probabilities, a reliable determination of these probabilities will need to be based upon a sufficiently weighty combination of evidential features. And just the same concern for the weight of relevant evidence would be needed if, instead of trying to *predict* whether their client will survive to the age of 65, the company were trying rather to *explain* why he has done so.

Philosophers sometimes invoke, in this kind of connection, Carnap's ([1950], p. 211) requirement of total evidence. For a conditional probability judgement to be applied as the major premiss of a prediction

or explanation, they say, it must be based on all the relevant evidence available. Or, if they prefer a relative-frequency theory of probability to a logical-relation one, they may invoke Hempel's ([1965], p. 397) requirement of maximal specificity for the reference class. But in practice such requirements can rarely, if ever, be satisfied. Even to know just the *available* evidence (physical, meteorological, geological, astrophysical, epidemiological, sociopolitical, *etc.*) bearing on a person's survival to 65 one would have to work away indefinitely, since what is not available to-day might, with sufficient effort, be made available to-morrow. And almost certainly, if one were thinking in terms of relative frequencies, one would soon be reduced to a reference-class of one member–the person himself–so that no statistical data could be compiled. What is important in practice, therefore, is for such a probability to have as much weight as is permitted by the methodology of estimation, by pertinent economic constraints, by the current state of scientific knowledge, by the nature of the subject-matter and by any other limits to enquiry. So at least comparative judgements of weight have to be made, and it may be worth while investigating whether the weights of different probabilities can also be ranked or measured. But we can afford to take rather less interest in theoretical ideals like the requirement of total evidence. In the real world we are occupied with what is better and what is worse, not with what is perfect. (Of course, Bayesian conditionalisation may be seen in Keynesian terms as a device for increasing weight. But it too does not issue in explicit comparisons, rankings or measurements of weight.)

An illuminating way to look at the matter is this. Suppose that the estimated probability that a person will survive to age 65, on the evidence that he or she is a lorry-driver, is 0.8. This is a generalised judgement of conditional probability, in the sense that it does not assert anything about a particular person. From it we can derive its instantiation for a particular person only on the assumption that our reference to that person does not add to or subtract from the evidence on which the probability is conditional. (Indeed without that assumption we could hardly derive its instantiation as we can[1]–in counterfactual cases, *i.e.* even in the case of a person who is not in fact a lorry-driver.) But on the unconditional issue no such assumption is possible. If we want to infer the *unconditional* survival prospects of a particular person–say, Mr Smith–we cannot avoid the need to allow for the fact that Mr Smith may be specially circumstanced in some relevant way: lorry-driving may be his work, but hang-gliding may be his hobby. So inferability to a singular judgement of unconditional probability–to the judgement, say, that the probability of Mr Smith's surviving to 65, is 0.8–must depend on how

much of the relevant facts about Mr Smith is included in the premisses. That is, this inferability varies directly with the weight of the evidence. If $p(H/E) = n$, then the weight of $p(H/E)$ determines the strength of our entitlement to infer from E to $p(H) = n$.

At any rate it is convenient to speak thus in the present context. But strictly speaking, and to avoid any possibility of confusion, we should note that, because of the considerations just mentioned, there is some equivocation here between '$p(H) = n$' and '$p(H/E) = n$'. In the latter–the singularised judgement of conditional probability–each occurrence of the referring expression 'Mr Smith' may be replaced *salva veritate* by an occurrence of, say 'Mr Brown', whereas in the singularised judgement of unconditional probability that is not so. Correspondingly the value of this $p(H)$ is not related to the value of this $p(H/E)$ by Bayes' theorem. For, if '$p(H) = n$' had instead been used here to state a prior probability relative to '$p(H/E) = n$', we should have had to grant the former the same substitutivity entitlements as the latter. Indeed philosophers who use the expression 'statistical syllogism' to name an inference of the kind in question, from 'E' and '$p(H/E) = n$' to '$p(H) = n$', are speaking rather misleadingly. Genuine syllogisms, whether valid or invalid, do not equivocate thus between premiss and conclusion.

2. WHY DID KEYNES NOT APPRECIATE HOW WEIGHT MATTERS?

When we view the problem in the above terms it is easy enough to see not only how weight matters but also why Keynes was unable to appreciate how it matters. Because he held that probability should be thought of as a logical relation (and because he apparently had no conception of deducibility from the null premiss), he could hardly allow a place for judgements of unconditional probability other than as mere ellipses of 'ordinary speech' ([1921], p. 7). *A fortiori* he could hardly appreciate the existence of the problem of how to grade our entitlement to detach an unconditional probability from a conditional one. It is scarcely surprising therefore that Keynes admitted to doubting whether the theory of weight had any practical significance, since by his very treatment of probability as a function of ordered pairs of propositions he had cut himself off from the possibility of articulating the nature of this significance.

Nevertheless it is very much to Keynes's credit that he did not suppress his various intuitions about the nature of weight. And perhaps these intuitions were strengthened by the fact that, like most people, he could not altogether rid himself of the intuitive idea that probabil-

ity can also be conceived as the relative frequency with which one set shares its membership with another, while he recognised that the weight of a relative frequency can be increased by a relevant partitioning of the reference class ([1921], p. 27). Unfortunately, however, the relative-frequency theory of probability has a parallel difficulty in articulating just how weight matters. Just as the logical relation theory seems not to assign probabilities to single propositions (as distinct from ordered pairs of propositions), so too the relative frequency theory, seems not to assign them to single sets (as distinct from ordered pairs of sets).

3. IS WEIGHT INCREASED BY EACH ADDITION OF RELEVANT EVIDENCE?

In the standard, probability-theoretic sense of 'relevant' E_2 is relevant to $p(H/E_1)$ if and only if $p(H/E_1 \ \& \ E_2) \neq p(H/E_1)$. But, if the weight of an argument is to continue to grow with each increment of relevant evidence, it must clearly also be taken to grow even with the addition of a certain kind of irrelevant evidence. This is because one might add the evidence $E_2 \ \& \ E_3$ where E_2 alters the probability of the argument in one direction exactly as much as E_3 alters it in the other. In such a case the addition of E_2 on its own would increase the weight of the argument first, and then the addition of E_3 to the evidence would increase the weight yet further. So, since it should presumably make no difference whether E_2 and E_3 are added successively or conjunctively, the addition of the conjunction $E_2 \ \& \ E_3$ must also increase the weight of $p(H/E_1)$ even though it is not–in the standard sense–a relevant piece of evidence. Keynes therefore defined a proposition as 'relevant' to $p(H/E_1)$ in this connection if and only if it entails a proposition E_2 such that $p(H/E_1 \ \& \ E_2) \neq p(H/E_1)$. But it would obviously be simpler to retain the normal probability-theoretic sense of 'relevant', and then to say–using Keynes's functor 'V(.../- - -)' for 'the weight of the probability of..., given - - -'–that $V(H/E_1 \& E2) > V(H/E_1)$ if and only if E_2 entails a proposition that is relevant to $p(H/E_1)$.

Unfortunately this will not quite do as it stands. There is a further difficulty, which apparently Keynes did not see. According to classical logic any proposition E_2 you like, however irrelevant to $p(H/E)$, nevertheless entails the disjunction $E_2 \vee H$, and $E_2 \vee H$ is certainly relevant to $p(H/E_1)$ because $p(E_2 \vee H/H \& E_1) = 1$ and so by Bayes's theorem

$$p(H/E_1 \& (E_2 \vee H)) = \frac{p(H/E_1) \times p(E_2 \vee H/H \& E_1)}{p(E_2 \vee H/E_1)} = \frac{p(H/E_1)}{p(E_2 \vee H/E_1)}$$

where $p(E_2 \vee H/E_1) > 0$. So Keynes seems to have trivialised the concept of weight by allowing any proposition to increase the weight of

any argument. In order to avoid such trivialisation we need to tighten the conditions under which $V(H/E_1\&E2) > V(H/E_1)$. We need to say that this inequality holds if and only if E_2 entails a proposition E_3 that is relevant to $p(H/E_1)$, where no proposition E_4 occurs in E_3 (or in any equivalent of E_3) such that E_2 entails E_4 and, without affecting the relevance of E_3 to $p(H/E_1)$, E_4 can be replaced in E_3 (or in some equivalent of E_3) by a proposition that has no relevance to $p(H/E_1)$. And we can also say that under just these same conditions E_2 will give at least as much weight to $p(H/E_1)$ as E_3 does.

4. DO ARGUMENTS INHERIT WEIGHT VIA THE ENTAILMENTS OF THEIR CONCLUSIONS?

We have seen in **3** how entailment between evidential propositions affects weight. The question also arises whether entailment between conclusions carries with it any necessary consequences for assessments of weight. But the answer must be that it does not. For we know that, where H_1 entails H_2, E_2 may be relevant to H_2 on E_1 and yet not relevant to H_1 on E_1, or to H_1 on E_1 and yet not relevant to H_2 on E_1 ([Carnap], 1950, pp. 348-97). On the other hand it is certainly reasonable to assume, so far as weight is a grading of inferability, that the weight of an argument is unaffected when any proposition in its premisses or conclusion is replaced by another proposition that is necessarily equivalent to it.

5. CAN ARGUMENTS IN CO-ORDINATE TERMS BE COMPARED FOR WEIGHT?

The question now arises whether any comparisons of weight can be drawn between $p(H_1/E_1)$ and $p(H_2/E_2)$ when no entailments hold in either direction between E_1 and E_2. It is tempting to claim, for example, that $p(H_1/E_1)$ and $p(H_2/E_2)$ should be said to have the same weight when each pair of propositions ascribes the same pair of predicates though to different individuals, and Keynes does claim this ([1921], p. 73). After all, the weight of a probabilificatory argument ought surely to remain invariant under uniform exchanges of the individuals to which it refers, much as the validity of a causal or logical argument remains invariant under these conditions. But the same point is perhaps more accurately made by treating weight as being primarily a property of generalised conditional probabilities, and derivatively of their substitution-instances. Indeed, the weight of the evidence is the same for a generalised conditional probability as for each of its substitution-instances because

the relevance of the evidence is the same (on the assumption that our reference to the instantiating individual does not add to or subtract from the evidence on which the probability is conditional).

It is also possible to extend the theory of weight by assuming that the weight of an argument is unaffected when each occurrence of a particular predicate P_1 within its premisses or conclusion is replaced by an occurrence of another predicate P_2, just so long as P_1 and P_2 are both members of the same family of co-ordinate but mutually inconsistent predicates. A predicate might thus derive its weight-increasing potential for a given argument from the relevance to that argument of another predicate in the same family, or from its relevance to an argument formulated in terms of predicates that belong to the same families as the predicates in the given argument.

What would be the justification for this? First, it seems reasonable that, when we have conducted an enquiry to find out whether a candidate for life insurance runs exceptional health risks in his leisure-time activity, we should be able to add the results of the enquiry to the premisses of the argument for his survival to 65 in such a way as to give that argument the same increment of weight independently of whether the probability of the argument is affected by this (as might be the case if his only hobby were discovered to be hang-gliding) or not affected (as might be the case if his only hobby were discovered to be stamp-collecting). After all, whichever the outcome the detachment of a probability for the argument's conclusion is now protected in relation to that issue, and the primary purpose of assessing weight is, as we have seen, in **1**, to evaluate such detachability. So though Keynes's test gives us no authority for the idea, it is difficult to avoid supposing that a viable theory of weight should allow invariance of weight under changes of a premiss's predicate within the same family: if one member of the family increases the weight of an argument to a particular conclusion, each of the others does also, and if one argument to a particular conclusion with premisses $E_1, E_2 \ldots E_n$ has its weight increased by a certain additional premiss then so does every other argument (to the same conclusion) that differs only by replacing a predicate in one of $E_1, E_2 \ldots E_n$ by a term co-ordinate with it.

Secondly, if the value of the probability $p(H_1/E_1)$ is known to be affected by the addition of E_2 to the premisses, a change is thereby made in the constraints that are known to affect the value of any probability $p(H_2/E_1\&H_2)$ where H_2 is inconsistent with H_1, because we know mathematically that $p(H_2/E_1\&E_2) \leq 1 - p(H_1/E_1\&E_2)$. So it is reasonable to suppose that E_2 increases the weight of $p(H_2/E_1)$ just as much as it increases that of $p(H_1/E_1)$. Keynes speaks of this only in

the case in which H_2 is the contradictory of H_1, where he accepts that $V(H/E) = V(\overline{H}/E)$. But the principle at issue is a more general one, extending to families of mutually exclusive predicates as well as to pairs of contradictory ones.

We see thus that the same principle is at work in regard to both the premisses and the conclusion of an argument. The weight of the argument is unaffected by the substitution of one predicate for another within the same family of co-ordinate but mutually exclusive predicates.

6. DOES WEIGHT HAVE ANY LIMITING CASES?

If a necessarily true proposition is added to an argument's premisses (whatever they may be) it cannot change the probability of the argument. Hence it adds no weight to an argument, and if all the premisses of an argument are necessarily true, the argument has minimal weight.

On the other hand, according to classical logic, if a necessarily false proposition is added to an argument's premisses, it entails every proposition and a fortiori therefore it entails every proposition that is relevant to the conclusion. So on this basis, if the premisses of an argument contain a contradiction, the argument has maximal weight.

If an argument has a necessarily true conclusion, it has a probability of 1 whatever the premisses. So no additional premiss can affect its probability, and it must therefore be regarded as already having maximal weight. (Correspondingly, if H is necessarily true, we are fully entitled to detach $p(H) = 1$ from $p(H/E) = 1$ whatever E may be.)

Similarly if an argument has a necessarily false conclusion, it has a probability of 0 on any premiss, and again no additional premiss can affect its probability. So it too has maximal weight (as, indeed, we could also show by combining the principle $V(H/E) = V(\overline{H}/E)$ from **5** with the principle established in the previous paragraph).

Also, if the premisses of an argument already entail its conclusion, or if they already contradict its conclusion, no additional premiss can affect its probability. So under these conditions too the argument has maximal weight.

7. IF ONE PREMISS IS OF GREATER RELEVANCE THAN ANOTHER, DOES IT ADD MORE WEIGHT?

The question here–not discussed by Keynes–is this. Suppose $p(H/E_1\&E_2)$ differs more from $p(H/E_1)$ than does $p(H/E_1\&E_3)$. We can then say that E_2 has a relevance to $p(H/E_1)$ that is of greater extent

than E_3 has. Is the weight of $p(H/E_1\&E_2)$ therefore greater than that of $p(H/E_1\&E_3)$, or not? There are some reasons for saying that it is, but stronger reasons for saying that it isn't.

On the one hand it seems at first sight intuitively implausible to hold that the probability of a male person's surviving to age 65 has just as much weight as the probability that a person with a dangerous hobby will survive to age 65. If time or other resources for enquiry were in short supply, would it not be much more important in assessing a life insurance premium to determine whether the person concerned had a dangerous hobby than to determine what the person's sex was? And, in general, if we consider the kind of purpose for which comparisons of weight are needed, it looks at first as though the extent of a new premiss's relevance ought to enter into comparisons of incremental weight. If we want to detach a value for the probability of H, we seem better off starting with a value for $p(H/E_1\&E_2)$ than with one for $p(H/E_1\&E_3)$ if E_2 is more relevant to $p(H/E_1)$ than E_3 is.

On the other hand this way of assessing weight soon plunges into paradox. Suppose a set of evidential items $E_1, E_2, \ldots, E_{100}$ in regard to a hypothesised conclusion H. Suppose too that quite a lot of these items, on their own, ground low probabilities in favour of H, quite a lot ground high probabilities in favour of H, and quite a lot ground intermediate probabilities at varying levels. One way of ordering these items would be to begin with those highly in favour of H, then proceed with those slightly less in favour and so on down, ending up with those highly in favour of H. In such a carefully graduated order the extent of the relevance of each new piece of evidence, after the first, would tend to be small. So if the weight of the argument were to be affected by the extent of the relevance of each incremental piece of evidence, as well as by the number of those pieces, the additional effect on the overall weight would be minimal. But, if instead the evidential premisses were ordered so as to alternate as violently as possible between favourable and unfavourable items, the overall effect on the weight would be very different, if extent of relevance was allowed to affect the issue at each incremental step. Hence, if we allow the extent of an added premiss's relevance to affect the cumulative weight of an argument, we could end up with different weights for an argument to the same conclusion from logically equivalent premisses, just because we calculated the weights on the basis of different orderings for the addition of new premisses. This is certainly inconsistent with what was said about the substitutivity of equivalents in **4**, and seems unacceptably paradoxical. The total weight of the evidence for a conclusion ought to be independent of the order in which different evidential items are stated. For some people the psychological effect

may vary with the order of statement, but a competent reasoner ought to be able to discount such effects. So I conclude that in a theory developing Keynes's seminal idea any weight added to an argument by a new premiss should stem solely from the fact that this premiss entails a relevant proposition and not at all from the extent of that relevance.

8. IS WEIGHT DETERMINED BY RELATED PROBABILITIES?

The term 'weight' has been used by philosophers to cover one or other of a variety of probability-related measures. Thus Reichenbach ([1949], P. 465) held that, if we know the limit towards which the relative frequency of a certain kind of outcome tends within a sequence of events, then this value can be regarded as 'the weight of an individual posit concerning an unknown element of the sequence'. Or, in other words, 'the weight may be identified with the probability of the single case'. Clearly this is not the sense in which Keynes was using the term in chapter VI of his [1921], since he emphasised that an argument of high weight is not, as such, 'more likely to be right' than one of low weight. It is easy to find arguments that have high probability but low weight, or low probability but high weight. (Elsewhere in his [1921], however, Keynes sometimes uses the term 'weight' more loosely: *e.g.* p. 218.)

Again, Good [1968] has discussed weight in the sense in which the weight of evidence concerning H that is provided by E, given G, is equal to $\log \{p(E/HG) \div p(E/\overline{H}G)\}$. But the quantity of evidence relevant to a certain argument is independent of the probability of the evidence given the conclusion. A great quantity of evidence might have been collected in a murder trial, with most of it tending to incriminate the accused, but it might also include an unshakable alibi. In such a case the evidence available might have relatively low probability, given the innocence of the accused, but it would have a heavy Keynesian weight. Similarly, that a person died before the age of 80 given that he died before the age of 8 has maximum probability, but the evidential fact that he died before the age of 80 gives a rather small increment of weight to any argument that he died before the age of 8.

It is sometimes suggested that, if the probability that a person assigns to his belief may be quantified in terms of the odds at which he would accept a bet on its truth, given specified evidence, then the weight of that evidence may be taken to be reflected in the amount that he is prepared to bet: he may be expected to be willing to put a larger sum at risk when there is more evidence from which to estimate appropriate odds. But other considerations also may affect our attitude towards the

size of a bet. Suppose that there is a great deal of evidence, and that this evidence suggests the appropriateness of very long odds. Would you really be willing to risk losing just as large a part of your fortune then as you would risk losing if the odds, on the same evidence, were much shorter?

It is also sometimes suggested that the weight of an argument may be taken to vary inversely with the mathematical expectation of gain from a search for further relevant evidence. But this suggestion is open to at least two cogent objections. First, in order to avoid begging the question the gain talked about must presumably be in some non-epistemic kind of utility. And this raises familiar problems about the evaluation of epistemic functions by reference to non-epistemic criteria, as discussed in e.g. Levi's [1984].

Secondly what are we to say when, for example, a vital eye-witness has died without ever disclosing what he saw? The expectation of any kind of gain from further research in that direction may then be zero, but the weight of the evidence about what actually happened is not increased because of the missing data. This is because the weight of the evidence obtained is being assessed by comparison with the supposed totality of relevant facts, not with the supposed totality of discoverable relevant facts. So, even if we had all the available evidence, our argument might still not have maximal weight. Sometimes, for example, the prosecution cannot prove guilt beyond reasonable doubt even though someone must have committed the crime in question.

9. CAN ONE ARGUMENT BE COMPARED FOR WEIGHT WITH ANOTHER IF ITS TERMS NEITHER ENTAIL NOR ARE CO-ORDINATE WITH THE OTHER'S TERMS AND NEITHER ARGUMENT IS A LIMITING-CASE?

We have seen that arguments may be compared with one another for weight where certain entailment relations hold between their premisses (as in **3**), or where their terms are co-ordinate and mutually exclusive (as in **5**), and that weight is maximised or minimised in certain special cases (**7**), but that weight is not determined by the sizes of associated probabilities (**8**). The question now arises whether any other principles bear on the determination of weight. In particular, since the extent of a premiss's relevance does not affect its incremental weight (**6**), should we treat this as a licence to suppose, by what I shall call 'the principle of equipollence', that the members of different families of predicates

enhance the weight of an argument equally when they enter relevantly into its premisses? For example, in **7** we have seen reason to reject the view that the probability that a person with a dangerous hobby will survive to age 65 has greater weight than the probability that a male person will survive to age 65. But does it therefore follow that these two probabilities should be attributed the same weight? Is the inequality to be rejected because no comparisons of this kind are possible, or just because the true comparison is one of equality? I shall argue for the former thesis, or–more exactly–for the thesis that comparisons of this kind are possible only on the basis of unacceptable assumptions.

Consider the two predicates 'has a dangerous hobby' and 'has a dangerous hobby and a weak heart'. We obviously cannot accept (as the principle of equipollence would seem to require) that the weight of evidence is unaffected by which of these two predicates is ascribed in the evidential proposition, since having a weak heart is relevant to whether a person survives to age 65 even on the condition that he or she is a lorry driver and has a dangerous hobby. So either we have to reject the principle of equipollence altogether, or we must restrict its application to primitive predicates in some appropriately tailored language-system. But the latter kind of move would introduce a substantial element of linguistic convention into the assessment of weight. The weight of an argument would depend not just on facts about probabilistic relevance but also on which predicates were chosen as primitive and therefore as having no non-trivial entailments. If, for example, 'has a dangerous hobby' and 'is male' were both treated as primitive predicates, the probability that a person who has a dangerous hobby will survive to age 65 would be assigned the same weight as the probability that a male person will survive to age 65; but, if one of the two predicates were treated as primitive and the other not, the two probabilities would have different weights. So, unless there is reason in a particular area of enquiry to suppose that the primitiveness or non-primitiveness of a predicate is unambiguously determined by the facts rather than by convention, it looks as though the principle of equipollence cannot be rescued.

It would certainly be very convenient if each of the primitive predicates that might enter relevantly into the premisses of any argument for a certain conclusion added an equal weight and all the relevant predicates that added least weight were primitive. The addition of a premiss containing any one such predicate could then be supposed to add one unit of weight to the argument. But this convenient state of affairs is most unlikely to obtain. The primitiveness or non-primitiveness of a predicate is a relatively a priori property of it that cannot usefully be made to depend on facts about relevance to a particular argument for a

particular conclusion. There is a fairly obvious reason for this. An issue that has no relevance for one argument may have considerable relevance for another, whether the arguments be for the same conclusion or for different ones, and the structure of a language–including the primitiveness or non-primitiveness of certain predicates–cannot usefully be geared to one particular series of arguments for one particular conclusion. If it were so geared, the strategy would be self-defeating because the range of comparisons permissible within the language would be excessively narrow.

10. CAN WEIGHT BE RANKED AS WELL AS COMPARED?

If the principle of equipollence is indefensible, there is no natural unit of weight and the prospects of any non-arbitrary system for measuring weight are very poor. But that does not exclude the possibility of having a principled system for ranking it, at least in relation to arguments about a given subject-matter, *i.e.* about conclusions that involve predicates of a given family. Such a system would require that the weight of any two arguments about the given subject-matter should be comparable, that superiority in weight should be a transitive relation, and that a level of least weight should be recognisable. And we should have such a system if we had an ordering for a certain set of families of evidential predicates and concerned ourselves only with arguments from premisses that contain just predicates belonging to the first family, or just those predicates plus predicates belonging to the second family, or just predicates from each of the first three families, and so on cumulatively. Moreover, when our list of such families was finite and determinate, we would also have a recognisable level of maximum weight and could tell how far short of it a particular argument fell. But how are predicate-families to be ordered appropriately?

We have to bear in mind that the primary purpose of the weight-judging enterprise is to grade our entitlement to detach a value n for the probability of H, when given premisses E and $p(H/E) = n$. And we have to bear in mind also that the time or other resources available for enquiry may be limited and that it is desirable for the probability of H that we detach to be as close as is possible within those limits to H's true probability. Indeed, because it may be the case that very few individuals in the population available for sampling possess certain complex combinations of characteristics, it may be impracticable to determine a probability that has more than a certain level of weight. For reasons such as these it will be important to give priority in the ordering

of predicate-families to those families that contain at least one predicate which is highly relevant in relation to the accepted prior probability of at least one conclusion in the given field. This is where the intuition about extent of relevance that was discounted in **6** comes into its own. The predicate-family containing a predicate of greater relevance to some conclusion in the given field than any other predicate-family contains should be placed first in the ordering, and so on down. But, of course, two or more predicate-families might tie at any stage, and if so we should either have to resort to some arbitrary criterion of priority in their case–say, an alphabetical one–or better, if we want to avoid any element of arbitrariness in constructing our system of ranking, we should combine such a set of two or more tying predicate-families into a single predicate-family that contains every possible combination of the predicates belonging to the tying predicate-families. With our predicate-families thus ordered we should not only be able to rank the weights of any arguments from premisses of the kind described in the previous paragraph: we should also be able to ensure that, even if the argument was not based on a conjunction of all potentially relevant premises and so its weight was not maximal, then at least it would be based on a conjunction of the most relevant premisses so that the probability detached was as near to the true one as it could be for that number of premisses.

Moreover, if on the basis of the same ordered set of predicate-families we can rank evidential weight for *conclusions* containing predicates that belong to two or more different families, these weight-rankings are obviously going to be comparable with one another. For example, if E adds weight to the probability that Mr Smith will survive to 65, given that he's a lorry-driver, then presumably E adds the same weight to the probability that Mr Smith will survive to 66 given that he's a lorry-driver, even though surviving to 65 and surviving to 66 are not incompatible with one another.

11. IS IT WORTH WHILE KNOWING THE WEIGHT OF AN ARGUMENT WITHOUT KNOWING ITS PROBABILITY?

Keynes defines his concept of weight in such a way that it is possible to know the relative weight of an argument without knowing its probability. Yet the point of the weight of an argument, as we have seen in **1** above, is in order to be able to grade our entitlement to detach an unconditional probability from a conditional one. We need to know the relative value of $V(H/E)$ in order to grade our entitlement to infer $p(H) = n$ from

the conjunction of E with $p(H/E) = n$. So, associated with the weight of an argument from E to H, there is also a value n such that we are directed by $p(H/E) = n$ towards the assignment of n to $p(H)$ when E is given. The weight of the argument grades our entitlement to proceed in that direction. Hence, though the size of the weight and the value of the probability are independent of one another, we need to know the direction towards which the argument is headed if we are to be able to use our knowledge of its weight.

12. WHAT IS THE CONNECTION BETWEEN KEYNESIAN WEIGHT AND BACONIAN LEGISIMILITUDE?

Elsewhere, in a development of Francis Bacon's seminal ideas about inductive reasoning (Cohen [1970] and [1977]) I have argued for a method of ranking the reliability of any generalised conditional (or of any of its substitution-instances) within a particular field of factual enquiry by reference to the complexity of the controlled experiments that it survives. I call this 'the method of relevant variables'; and I call the parameter that it grades 'legisimilitude' (Cohen [1980a] and [1985]) i.e. proximity to being a natural law. Experiments of different degrees of complexity are (if properly insulated from external factors) like simulations of possible worlds that differ from one another in the variety of combinations of inductively relevant circumstances that they contain, and generalisations are shown to possess greater and greater legisimilitude as they are shown to hold good over varieties of possible worlds that are more and more richly stocked with combinations of inductively relevant circumstances. One can show too that these rankings of legisimilitude constrain one another in accordance with the principles of a modal logic that generalise on the Lewis-Barcan system S4.[2] It follows that universally quantified conditionals (or their substitution-instances) which are qualified in such a way as to withstand falsification by any experiment of a certain level of complexity may be attributed an appropriate level of legisimilitude by the same system of ranking.

The various lines of reasoning that I have developed in the past in order to justify this system of ranking Baconian legisimilitude and the lines of reasoning developed in **3-9** above in regard to the comparing or ranking of Keynesian weight are quite independent of one another. The Baconian scheme was defended in a context of concern with generalisations that are rooted in causality and it treats of generalisations about probabilities only as a special case. The Keynesian scheme has been defended in a context of concern with probabilistic relevance and only

a limiting case of it (where the probability involved is 1) instantiates deterministic generalisation. But both lines of reasoning converge on precisely the same underlying structure. The method of ranking weight that was defended in **9-10** is an application of the method of relevant variables, and the various logical constraints on the comparison or assignment of weight that were defended in **3-6** are all derivable within the logical syntax of legisimilitude.[3] More exactly, on a proper reckoning the Keynesian weight of $p(H/E)$, where $p(H/E) = n$, should turn out equal to the Baconian legisimilitude of $E \rightarrow p(H) = n$.

The importance of this fact seems to me to be that, even if your intuitions about causality are insufficiently powerful to drive you down the Baconian road, you may well find that your intuitions about probability will drive you down the Keynesian road to the same destination as that to which the Baconian road would have led you. Moreover it should be clearer now how Baconian (i.e. weight-orientated) modes of reasoning are not intrinsically in any kind of conflict with probabilistic ones but can serve to complement them.

Finally it is worth considering Keynes's theory of weight in relation to his proposal of a probabilistic mode of assessment for Baconian induction, i.e. for inductive support that depends on the variety of relevant evidence. When Keynes talked about such induction in Part III of his [1921] he had in mind the supporting of deterministic generalisations like 'All A are B', which he calls 'universal inductions', rather than the supporting of probabilistic ones like 'All A have an n per cent probability of being B', which he calls 'inductive correlations' or 'statistical inductions'. If he had thought more about the latter in connection with his theory of weight, he could have used it to provide an appropriate mode of assessment for their support and he might then have sought to extend the theory of weight so as to cover deterministic generalisations also and produce a general theory of Baconian legisimilitude. But in fact when he briefly ([1921], pp. 409-12) has regard to considerations of weight in connection with statistical inductions he does not use the term 'weight' at all or refer back to his earlier discussion of the subject.[4]

Notes

1. The difference between counterfactualisable and non-counterfactualisable probabilities is discussed in (Cohen [1986], §17-19). A generalised conditional probability that is non-counterfactualisable has zero weight, since its value is the outcome of an accidental relationship and affords no basis for inference, in a particular case, to the value of an unconditional probability.

2. There are isomorphisms also with the Levi-Shackle theory of potential surprise: see (Cohen [1980b], pp. 64-66 and p. 171).

3. The development of this logical syntax was begun in Cohen ([1970], pp. 216-37) and continued in Cohen ([1977], pp. 229-40): compare theorem-schema 710 there with **3** here, 248 with **4**, 357 and 613 with **5**, and 703, 704, 707 and 728 with **6**.

4. I am grateful to the Editor for helpful comments on an earlier draft, and also to participants in a discussion at the Conference on Cambridge Philosophy of Science, 1850-1950, held in Bologna, Italy, during May 30-June 1, 1985.

References

Carnap, R. (1950). *Logical Foundations of Probability.* Routledge and Kegan Paul, London.

Cohen, L. (1985). The verisimilitude-legisimilitude duality. In *La Nature de la Verite Scientifique*, volume 27 of *Archives de l'Institut Internationale des Sciences Theoriques*, pages 55–66.

Cohen, L. J. (1970). *The Implications of Induction.* Methuen.

Cohen, L. J. (1977). *The Probable and the Provable.* Clarendon Press.

Cohen, L. J. (1980a). 'Comments and Replies on the Levi-Teller session' and 'What has Inductive Logic to do with Causality?'? In Cohen, L. and Hesse, M., editors, *Applications of Inductive Logic.* Clarendon Press.

Cohen, L. J. (1980b). What has science to do with truth? *Synthese*, 45:pp. 489–510.

Cohen, L. J. (1986). *The Dialogue of Reason.* Clarendon Press, Oxford.

Good, I. J. (1968). Corroboration, explanation, evolving probability, simplicity and a sharpened razor. *The British Journal for the Philosophy of Science*, 19:pp. 123–43.

Hempel, C. G. (1965). *Aspects of Scientific Enquiry, and Other Essays in the Philosophy of Science.* The Free Press.

Keynes, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

Levi, I. (1984). *Decisions and Revisions: Philosophical Essays on Knowledge and Value.* Cambridge University Press.

Reichenbach, H. (1979). *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability.* University of California Press, second edition 1971 edition.

18

SOME STEPS TOWARDS A GENERAL THEORY OF RELEVANCE

Abstract The classical analysis of relevance in probabilistic terms does not fit legal, moral or conversational relevance, and, though analysis in terms of a psychological model may fit conversational relevance, it certainly does not fit legal, moral or evidential relevance. It is important to notice here that some sentences are ambiguous between conversational and non-conversational relevance. But, if and only if R is relevant to a question Q, R is a reason, though not necessarily a complete or conclusive reason, for accepting or rejecting something as an answer to Q. Reasons of this kind are governed by appropriate covering laws or principled probabilities and a number of questions thus arise about the relationship between relevance and certain formal-logical properties.

1. THE NATURE OF THE PROBLEM

Enquiry into the topic of relevance has suffered much from those who have taken a part of the topic as the whole.

Thus, Keynes[1] and Carnap,[2] for example, dealt with relevance as a relationship of evidence to hypothesis. For them a specified additional piece of evidence was relevant to a specified hypothesis if and only if – roughly – the mathematical probability of the hypothesis on the existing evidence differed from its probability on that plus the additional evidence. And their detailed discussions of the topic were devoted to making the relatively minor improvements that they thought necessary in this classical analysis – improvements that involved supposing a somewhat more complex relationship between underlying mathematical

probabilities. For example, what is to be said about the relevance of the conjunction of I and J to H on existing E when I is positively relevant and J is negatively relevant to the same extent? Keynes and Carnap have been followed by others working in the same tradition and proposing even more sophisticated analyses.[3] But, even if this type of measure were an acceptable explication of evidential relevance in all its forms, it is hardly appropriate to moral or legal relevance, as when we say that a patient's own wishes are relevant to the moral legitimacy of certain types of medical treatment, or that in UK law a trader's country of residence is relevant to his tax liability. Certainly the moral legitimacy of a type of medical treatment is not to be equated with the probability of the treatment's being carried out, or with the probability of its being successful, nor is the legal liability of a person to pay a tax equatable with the probability that he will pay it.

Again, Sperber and Wilson say that "the value of the theoretical notion of relevance ultimately depends on the value of the psychological models which make use of it and especially on the value of the theory of verbal comprehension that it allows us to formulate".[4] But moral or legal relevance affords no special basis for psychological modeling. And an analysis of evidential relevance has to go beyond the theory of verbal comprehension: the footprints in the shrubbery may be relevant to the solution of the murder mystery without affecting, or being affected by, the comprehension of any linguistic communication. Nor is Grice's conversational maxim "Be relevant!"[5] readily translatable into the requirement that a speaker should aim to bring about an increase or decrease in an appropriate mathematical probability.

What is going on here? Obviously one possibility is that the expressions requiring analysis are a set of accidentally associated homonyms – 'relevance' in the sense of evidential relevance, 'relevance' in the sense of causal relevance, 'relevance' in the sense of moral relevance, 'relevance' in the sense of conversational relevance, and so on. But that type of solution is as intuitively implausible and theoretically unsatisfactory in the case of 'relevance' as Aristotle thought it to be in the case of 'goodness'.[6] It is to be adopted only if there is no unitary or systematic explanation to fit the phenomena. And in what follows I shall try to develop such an explanation. I shall argue that there is a concept of relevance that has an underlying pattern of interconnections which unifies it, despite the variety of criteria for applying the concept. But I know of no other way to try and establish this than to describe – with plausible examples – what seems intuitively to be the unitary underlying structure and to be prepared to modify this hypothesis if important counterexamples to it are produced.

2. CONVERSATIONAL RELEVANCE AND NON-CONVERSATIONAL RELEVANCE

An important ambiguity in many sentences about relevance is too often unrecognised or at least unacknowledged. The ambiguity is important here, because it might be argued that, if there are at least two radically different kinds of sentence by which many sentences about relevance may be paraphrased, then no unitary account of relevance is possible. Let us examine this ambiguity first in the light of a few examples. Once we have such a natural history of the phenomenon it will be easier to see whether a single analytical model can be constructed to fit the underlying variety of structure.

Consider, for example, the sentence

(1) The presence of footprints outside the window is relevant to any discussion whether the butler was the murderer.

On one reading (1) is about evidence – the kind of relevance with which Keynes seems concerned – and is roughly equivalent to

(2) The presence of footprints outside the window is evidence about whether the butler was the murderer.

But on another reading (1) is about conversational connection – the kind of relevance that Grice's maxim enjoins – and is roughly equivalent to

(3) Discourse about the presence of footprints outside the window would not be out of place in the context of a discussion about whether the butler was the murderer.

The sentences (2) and (3) are certainly not equivalent to one another because, though (2) states what is irrefutably a reason for accepting that (3) is true, (3) does not state what is irrefutably a reason for accepting that (2) is true. All concerned might know for example. that the footprints were made by the detective investigating the murder and therefore had no evidential relevance to questions about who was the murderer.

Again, consider the sentence

(4) In our conversation about early nineteenth century European history my description of Napoleon as a reformer is relevant to your describing Wellington as a reactionary.

On one reading (4) states an instance of causal relevance within a conversational exchange in which the description of Napoleon as a reformer preceded, and perhaps prompted, the description of Wellington as a reactionary, just as in, say,

(5) The insults that Peter experienced in the morning were a relevant factor in how he spoke in the afternoon.

But on another reading (4) states an instance of conversational relevance in which the description of Wellington as a reactionary preceded, and created an appropriate context for, the description of Napoleon as a reformer.

It is important to bear in mind these differences, because at first sight they seem to support a particularly tempting objection to the possibility of there being a single underlying structure in all types of discourse about relevance: if a sentence like (1) or (4) is genuinely ambiguous, surely there must be more than one underlying structure that is the key to its analysis?

We shall see later why this is not the case. But some more examples may help to clarify the underlying intuitions here. Consider the sentence

(6) George's revelation of the plot is relevant to Bill's decision to take the job.

On a non-conversational reading (6) is about the justifiability of a decision and roughly equivalent to

(7) George's revelation of the plot affects the merits of Bill's decision to take the job.

But on a conversational reading (6) is about conversational connection and is roughly equivalent to

(8) George's revelation of the plot would not be out of place in the context of Bill's consideration whether to take the job.

And the sentences (7) and (8) are apparently not equivalent to one another, because according to its most natural interpretation (7) points to an argument about what Bill ought to do, whereas according to its most natural interpretation (8) concerns the coherence of a piece of conversation.

Or consider the sentence

(9) The rule that no-one should be judge in his or her own cause is relevant to whether one of the local landowners should sit as a magistrate in the present case.

On a non-conversational reading of (9) it is about the content of a legal system and is roughly equivalent to

(10) The rule that no-one should be judge in his or her own cause is a principle relevant to the issue whether one of the local landowners should sit as a magistrate in the present case.

But on a conversational reading (9) concerns conversation about the content of a legal system and is roughly equivalent to

(11) Discourse about the rule that no-one should be judge in his or her own cause would not be out of place in the context of a discussion about whether one of the local landowners should sit as a magistrate in the present case.

And, again, the sentences (10) and (11) are not equivalent to one another. For, according to (10), the rule that no-one should be judge in his or her own cause is a principle that might tend to disqualify one of the local landowners from sitting as a magistrate in the present case, whereas according to (11) the rule that no-one should be judge in his or her own cause might just be one among many legal rules that could be appropriately mentioned in the specified context of discussion.

So the classical analysis – in terms of the difference between prior and posterior probability – could at best fit certain kinds of non-conversational relevance. Specifically, it might fit evidential or causal relevance. But it is altogether inappropriate for conversational relevance, because even where saying that X would be very appropriate after it had been said that Y the probability that Y was actually said might be very small. And yet any analysis that can help to supply a basis for a general theory of verbal comprehension must be concerned in part with conversational relevance. Conversely it is often easier for a speaker to maintain conversational relevance – as Grice's conversational rule instructs us to do – than to maintain non-conversational relevance. For we can often continue to speak about something when we can do nothing that might help to bring it about. So non-conversational relevance cannot be conceptually reduced to conversational relevance.

3. WHAT TYPE OF ENTITY DOES RELEVANCE CONCERN?

'Relevant' is clearly a relational term. What is its domain? Superficially there is an enormous variety of types of entity that in suitable contexts can be said to be relevant – conversationally or non-conversationally – to something. These include objects, actions, states, events, processes, properties, facts, rules, principles, assertions, commands, questions, issues, attitudes, and many other things. But it turns out that a statement about the relevance of an entity that is not a proposition can normally be reformulated as a roughly equivalent statement about the truth or consideration of a proposition appropriately related to that entity. For example, the sentence (2) can be reformulated as

(12) The truth of the proposition that there are footprints outside the window is relevant to the question whether the butler was the murderer,

and the sentence

(13) The length of the footprint is relevant to who is the murderer

can be reformulated as

(14) The truth of a proposition describing the length of the footprint is relevant to who is the murderer,

while the sentence (3) can be reformulated as

(15) Consideration of the proposition that there are footprints outside the window is relevant to consideration of the question whether the butler was the murderer,

where 'consideration' covers either spoken or non-spoken thought and either assertoric or non-assertoric attitudes.

There is also a wide variety of types of entity to which, in suitable contexts, something can be said to be relevant. Admittedly it is not easy to see how anything could usefully be said to be relevant to a physical object, such as a pen or a planet. But it does seem that, at least superficially, everything else that can be conversationally or nonconversationally relevant to anything also falls into the category of entities to which things can be conversationally or non-conversationally relevant. And again it turns out that a standard, normal form is discernible. Specifically, a statement about relevance to actions, states, events, processes, properties, facts, rules, principles, assertions, commands, attitudes, etc. can normally be reformulated as a roughly equivalent statement about relevance to a corresponding question or to consider ation of a corresponding question (where 'question' means 'issue' or 'problem' rather than some actually uttered interrogative sentence, but the question may have any interrogative structure). Thus nonconversational relevance relates the truth of a proposition, on the one side, to a question on the other, while conversational relevance relates the truth of a certain kind of proposition – one stating the occurrence of an utterance (and thus the consideration of a certain proposition – to consideration of a question on the other. For example, in accordance with whether (6) may be paraphrased by (7) or (8), it can be reformulated either as

(16) George's revealing the plot is relevant to any question about what are the merits of Bill's decision to take the job,

or as

(17) George's revealing the plot was relevant to Bill's consideration of the question whether to take the job.

Granted that relevance is thus at least a two-term relation, we may wish to ask whether a third term – a point of view – is always at least implicit. That is, we may wish to ask whether relevance must always be conceived as holding good only from a particular point of view.

It is obvious, however, that relativity to a particular point of view is not essential to non-conversational relevance. For, though it certainly makes sense to say

(18) The footprints are relevant for George, though not for Mary, to whether the butler was the murderer,

nevertheless this statement can be paraphrased without loss by

(19) George, though not Mary, thinks that the footprints are relevant to whether the butler was the murderer,

where 'relevant' clearly signifies a dyadic and supposedly objective relation. But even conversational relevance, as in (3), (8) or (11), is not essentially triadic and relative to a point of view. It is only when conversational and non-conversational relevance are confounded together that it may seem necessary to treat relevance as relative to a point of view. For then it may seem that the relevance of footprints outside the window, for example, concerns at the same time not only the question whether the butler was the murderer but also some person or person's discourse about this issue.

4. UNDER WHAT CONDITIONS IS ONE THING RELEVANT TO ANOTHER?

Non-conversational relevance, as we have seen, is a relation - typically appreciated or pointed out in the course of a reasoned enquiry, debate, conversation, meditation, explanation, or justification – between a true proposition and an askable question. More precisely, a true proposition R is non-conversationally relevant to an askable question Q if and only if there is a proposition A such that the truth of R is or would be some reason, though not necessarily a complete or conclusive reason, for anyone's accepting or rejecting A as an answer to Q, where this reason is warranted by some non-conversational principle. For example, (13), and therefore (14), are true if and only if there is a proposition A such that a true proposition describing the length of the footprint is or would be an evidentially based reason for anyone's accepting or rejecting A as an

answer to the question who is the murderer. Moreover a true proposition R is non-conversationally more relevant than S to Q if and only if there is a proposition A such that, for any proposition B, R is or would be a better reason for anyone's accepting or rejecting A as an answer to Q than S is or would be for anyone's accepting or rejecting B as such. And now conversational relevance can be seen to have the same underlying structure as non-conversational relevance, because one remark is relevant to another if and only if occurrence of the latter remark affords some conversational ground for inferring occurrence of the former. Or, more precisely, an utterance U is conversationally relevant to consideration of a question Q if and only if there is a proposition A such that the truth of the proposition that U occurred is or would be some reason, though not necessarily a complete or conclusive reason, for anyone's accepting or rejecting A as an answer to the question whether Q was – or will be - considered.

Note here, however, that the temporal order of thoughts corresponding to this implicit reasoning is not crucial to the difference between conversational and non-conversational relevance. Pointing out such a difference of temporal direction may sometimes help to clarify the difference between a conversational and a non-conversational reading of an ambiguous sentence, as in the case of (5), (6) and (7). But some kinds of remarks about conversational relevance involve reasoning forwards, as when a lecturer may want to warn his audience that what he has just been saying is also relevant to the issue that he will shortly be discussing. And some kinds of non-conversational relevance involve reasoning backwards, as in using present evidence to investigate past events: how the butler responds to to-day's questioning is relevant to whether he committed yesterday's murder.

What counts as a reason, or a better reason, in the analysis of relevance is obviously going to vary widely between different types of context. Thus, what is morally relevant is determined by the criteria for valid moral reasoning. For example, the patient's own wishes, if knowable, are relevant to whether a form of medical treatment that has dangerous side-effects may be used on the patient, because these wishes, if knowable, are normally a reason for, or against, the moral legitimacy of using the treatment. What sanctions there being such a reason here is a particular covering law that helps to license inferences in normal circumstances from appropriate kinds of knowable facts to conclusions about what is morally right or wrong in medical practice. What is legally relevant, within a particular system of jurisprudence, is determined analogously by the rules of the system. Moreover in both moral and legal contexts objections may be raised against claims to some stated factor's

relevance on the grounds that the supposed covering law has to be rejected. And – other things being equal – the more the qualifications that have to be conjoined together in the antecedent of the covering law in order to license the required inference conclusively in the combination of circumstances concerned, the less relevant is any one of those circumstances to the issue, because any one such circumstance is thereby rendered less conclusive a reason.

Evidential relevance is often similarly structured. The covering law may be a supposed law of nature, causal uniformity or standard pattern of human behaviour. Thus the state of the barometer to-day may be a reason for accepting that it will, or will not, rain to-morrow, and a proposition about the former state of affairs is therefore relevant to a question about the latter prediction. And again the supposed covering law may have to be rejected in certain circumstances, while the degree of relevance is measured inversely by the extent of the qualification that the covering law requires to incorporate into its antecedent in order to license a successful prediction in the combination of circumstances presented. The state of the barometer is more relevant to to-morrow's weather if a falling barometer is an infallible sign, or precursor, of rain to-morrow than if a falling barometer predicts this only when the wind is in the west.

Similarly, as a case of moral relevance, (7) will be true if and only if there is a proposition A such that the truth of the proposition about George's revelation is some reason based on moral laws (though not necessarily a complete or conclusive reason) for accepting or rejecting A as the answer to a question about the merits of Bill's decision.

Of course, the underlying structure of a sentence about relevance may not always be quite so simple. Consider, for example,

(20) The age of the patient is a relevant variable for testing the efficacy of a medicine.

The underlying issue here is obviously

(21) How should the efficacy of a medicine be tested?

But because of the level of generality at which relevance is being asserted in (20) it is not so obvious how an analysis of (20) can cohere with the analyses of particularising propositions like (1)-(19). The situation is clarified, however, when appropriate quantifiers are introduced into the analysans, as in

(22) For any medicine M, and any sufficiently large set of testable individuals T, there is a proposition A about M and T such that a true proposition describing the ages of the members of T is some

reason (i.e., not necessarily a complete or conclusive reason) for anyone's accepting or rejecting A as an answer to the question: how should the efficacy of M be tested?

The evidential or causal relevance of R to A may also be inferred from there being a difference between the causally based prior probability of A and the causally based probability of A given R, as when heavy smoking is said to be relevant to the expectation of lung cancer. The relevance arises because, given the probabilities, we infer that a man's heavy smoking is a reason, though not necessarily a conclusive reason, for expecting him to develop lung cancer.[7] But it would be a mistake to suppose that this type of relevance, familiar though it may be, somehow lies at the heart of the matter. Respect for the difference between causally based prior probability and causally based posterior probability is just one criterion of relevance alongside the others – that is, alongside conformity to moral principles, legal rules, laws of nature, rules of conversation, and so on. Anything that can sanction a reason, even if an incomplete or inconclusive reason, for accepting a particular type of proposition as a correct answer, or for rejecting it as an incorrect answer, to an askable type of question can count as a criterion of relevance.

It needs to be emphasised here, however, that the covering laws that can generate reasons in this way are not just accidentally true generalisations, but generalisations that are principled and thus distinguishable by their ability to license the deduction of counterfactual conditionals. That is because we want to be able to say not only that George's heavy smoking is relevant to his expectation of lung cancer, but also that, if George had been a heavy smoker, it would have been relevant to his expectation of lung cancer. Analogously, where the existence of such a reason stems from the difference between a prior and a posterior probability, these probabilities are causally based and therefore counterfactualisable. That is to say, if any actual member of the reference-class has a specified probability of having a certain characteristic, then, if some other entity had been a member of the class, it too would have had that probability. Being a heavy smoker would still have been relevant to a person's expectation of lung cancer even if more people had been heavy smokers. Contrariwise if, say, twenty people happen to be passengers on the same bus and five happen to have surnames beginning with the letter C, then the 0.25 probability that anyone picked at random from the passengers will have such a surname is non-counterfactualisable insofar as no underlying principle is assumed to be at work ensuring that one out of every four additional passengers would tend to have such a surname.[8]

The reasons at stake in conversational relevance (see Section 2) must be analysed analogously in terms of appropriate principles, whether these be thought of as psychological covering laws (co-ordinate with, say, meteorological or pharmacological laws), or as rules of pragmatics (co-ordinate with, say, rules of semantics or syntax). And it may be tempting to object that, despite what has been argued here about the connection between relevance and reasons, nothing has been said to exclude – overall – a naturalistic account of relevance. Specifically someone may argue that, if your reasons for accepting a particular proposition were the facts that caused your acceptance, then each of these facts was relevant to the question that this proposition answered. All kinds of relevance will then depend altogether on causation, and in this respect naturalised epistemology will have triumphed.

But relevance, like reasoning, has a prospective dimension as well as a retrospective one. It helps prediction as well as explanation. In the normal kind of retrospective case it is indeed tempting to identify the reason why someone did something with the cause of his so doing. But, in one familiar sense of 'reason', this identification cannot be correct, because in that sense the reason why someone did something may be the very reason that was given him – in advance – to do it, and in advance of the action the reason put forward for doing an action cannot at the same time be regarded as the cause of its actually being done.

When someone tells you a reason for accepting a particular proposition, for example, the validity of the reason does not depend on the success of the argument. The reason may be a valid one even if you yourself fail to be convinced, or it may be an invalid one even though it in fact convinces you. Correspondingly you may learn the truth of a proposition that is relevant to your enquiry without recognising that it is relevant, or you may think it relevant even though it is not. Sometimes, certainly, the occurrence of one speech-act or mental event may be said to be relevant to the occurrence of another in a context in which the relevance arises out of an underlying causal connection. Your shouted warning may have been a relevant factor in my deciding not to jump. But this can always be conceived as a kind of way in which the occurrence of one event may be non-conversationally relevant to the occurrence of another. It does not have to be regarded as a kind of way in which conversational relevance submits to a naturalistic account.

5. HOW DOES DEDUCTIVE LOGIC AFFECT RELEVANCE

A number of questions arise consequentially about the relationship between relevance and certain formal-logical properties. I shall here mention only four of these.

First, it is obvious that a logical law can sometimes be a covering law that generates relevance. For example, the law *modus ponendo ponens* makes the conjunction of two true propositions B and 'If B is true, then A is true' relevant to a question Q if A is an answer to Q: the truth of the conjunction is a reason for accepting A as an answer to Q. But there are some systems of deductive logic in which not all logical laws generate relevance. In a system of strict implication, for example, there is a law that, if A is necessarily true, then for any proposition B, A is strictly implied by B. But that does not make it at all plausible in such a case to hold, for a particular B, that the truth of B is a reason for accepting A as an answer to some askable question Q. Correspondingly B would not necessarily be relevant to Q. For example, the fact that it is raining is not relevant to the question whether 29 is a prime number. And in general we must expect that intuitive judgements of deductive-logical relevance and irrelevance will reflect intuitive judgements about what is a deductive-logical reason for what, while projects for the formal reconstruction of the former type of judgement will face all the difficulties encountered by projects for the formal reconstruction of the latter type.

You may object that this puts the cart before the horse. The formal analysis of human reasoning, you may say, must exploit a pre-existing concept of relevance, as the work of Anderson and Belnap[9] has shown: one cannot hope instead to clarify the concept of relevance in terms of the concept of a reason.

However the trouble is that the intuitive conception of relevance exploited by Anderson and Belnap is too specialised for present purposes. Thus, they hold it to be a necessary condition of A's entailing B, in the pure calculus of entailment, that A should be relevant to B. And, in their sense, a necessary condition for relevance in such cases is that A and B should share at least one propositional variable. So in applications of the calculus an entailing proposition must have some element of non-logical meaning in common with the entailed one. But clearly no such condition need hold in the case of reasons that are sanctioned by material principles rather than deductive-logical ones, as when the occurrence of the Big Bang is said to be relevant to the existence of supernovas. Nor can the objector fall back here on Anderson and Belnap's other criterion of relevance – the criterion requiring that for A to be

relevant to B it must be possible to use A in a deduction of B from A. For in applying this criterion a prior concept of deductive reasoning is explicitly presupposed.

Secondly, it is interesting to examine how relevance is related to extensionality. What emerges is that, where the propositions involved do not themselves contain any contexts that are resistant to the substitutivity of identicals, the statement that one such proposition is relevant to another does not introduce any resistance to substitutivity. For example, the fact that Scott wrote *Waverley* and that the author of *Waverley* is the author of *Ivanhoe* is indubitably relevant to the question whether Scott wrote *Ivanhoe*. And it is reasonable to feel that all judgements of relevance, whatever their content, are fully extensional. The fact that the author of *Waverley* is the author of *Ivanhoe* and that George III believed that Scott wrote *Waverley* may still seem nonconversationally relevant to the question whether George III believes that Scott wrote *Ivanhoe*, even if that fact does not entail an affirmative answer to the question. There seems to be some reason for accepting the proposition in question even if not a conclusive one.

Thirdly, it is clear that relevance is neither transitive nor intransitive. Transitivity is ruled out by the fact that A may be relevant to the conjunction of B and C (because it is relevant to B) and the conjunction of B and C may be relevant to D (because C is relevant to D) without A's being relevant to D. And intransitivity is ruled out by the fact that, if B is a simple – uncompounded – fact, and A is, say, causally relevant to B and B to C, then A is causally relevant to C. Fourthly relevance is neither symmetric nor asymmetric. It is not symmetric (and therefore also not reflexive) because, for example, it may be the case that A is causally relevant to B without B's being causally relevant to A. And it is not asymmetric because in the course of a conversation, say, two remarks may be relevant to one another.

6. TO WHAT IS AN ANALYSIS OF RELEVANCE RELEVANT?

Much, if not all, of what I have been saying may seem somewhat simplistic. One could certainly add a lot of fine print about erotetic logic, for example, if all relevance has to be relevance to a question. One could also enlarge at length on the different ways in which different kinds of statement about relevance must be handled. But even the broadest outlines of a general theory of relevance may nevertheless be worth sketching, because there has long been a tendency for philosophers to use the concept of relevance in certain special contexts – for example,

in the discussion of Bayesian confirmation, of Andersonian entailment, or of Gricean implicature – as if the criteria operative in this or that special context reveal the true nature of relevance. Instead of pursuing that kind of bottom-up project I have tried instead here to adopt a top-down approach, much as one might wish to clarify the general sense of the term 'good' before examining the criteria for using it in particular contexts. Indeed, it is only when the general sense of 'relevance' has been adequately analysed that one can begin to see why the criteria operative in particular contexts are appropriate to those contexts.

Notes

1. Keynes, 1921
2. Carnap, 1950
3. For example, Gardenfors, 1978 and Schlesinger,
4. Sperber and Wilson, 1986, pp. 119-20.
5. Grice, 1975, pp. 64-75.
6. Aristotle, 1890, p. 8
7. I leave aside because they do not concern the present issue the problems that arise about how precisely to measure relevance when it is thus inferred.
8. In regard to counterfactualisable judgements of probability see Cohen, 1986,pp. 165-166.
9. Anderson and Belnap, 1975, pp. 32-33

References

Anderson, J. R. and Belnap, N. (1975). *Entailment: the Logic of Relevance and Necessity, Vol. 1.* Princeton University Press, Princeton, New Jersey.

Aristotle (1890). *Nicomachean Ethics.* Clarendon Press, Oxford.

Carnap, R. (1950). *Logical Foundations of Probability.* Routledge and Kegan Paul, London.

Cohen, L. J. (1986). *The Dialogue of Reason.* Clarendon Press, Oxford.

Gardenfors, P. (1978). On the logic of relevance. *Synthese*, 37.

Grice, H. P. (1975). Logic and conversation. In Davidson, D. and Harman, G., editors, *The Logic of Grammar.* Dickinson, Encino.

Keynes, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

Schlesinger, G. N. Relevance. *Theoria*, 52.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and Cognition.* Blackwell, Oxford.

19

SHOULD A JURY SAY WHAT IT BELIEVES OR WHAT IT ACCEPTS?

1. Is Cognition Active Or Passive?

There is an old dispute in Western philosophy about whether the human mind is active or passive in the process of acquiring knowledge concerning the facts and laws of nature. According to Descartes, we articulate such knowledge in voluntary judgements whereby we either assent to or dissent from some relevant mental representation.[1] Thus each item of conscious knowledge is gained through an act of free will. But according to Hume we learn about nature via the formation of beliefs and, on his view, a person's beliefs are not controllable by willpower.[2] In other words on Hume's view our knowledge of nature is not acquired by voluntary acts of cognitive judgment, as Descartes held, but by the involuntary growth of cognitive feelings.

Reasons exist for holding that this old issue is still an important one, although some of the terminology in which it needs to be formulated may have changed. Indeed there is quite a range of currently discussed philosophical problems in relation to which we can profitably distinguish between active and passive cognitions, so as to clarify the different but interconnected roles that such cognitions play in human thinking. In the present paper, after some background remarks about how this distinction operates, I want to apply the distinction to the particular case of courtroom verdicts on matters of fact: should a verdict declare what a jury passively believes or what it actively accepts?

2. How Are Belief And Acceptance Related To One Another?

It is convenient initially to pick out a pair of expressions that will cover the fundamental difference at stake. Let me use the words *belief* and *acceptance* for this purpose. Where the variable letter 'p' stands for any proposition, 'belief that p'–in my sense as in the everyday one–is a disposition, when attending to issues raised or items referred to by the proposition that p, normally to *feel* it true that p and false that not-p, whether or not one is willing to act, speak, or reason accordingly. But to 'accept that p'–in the relevant sense of accept–is to treat it as given that p. It is to have or adopt a policy of deeming, positing, or postulating that p–that is, of going along with that proposition in one's mind as a premise or license for inference in some or all contexts for one's own and others' deductions, inductions, proofs, argumentations, inferences, deliberations, and so on, whether or not one feels it to be true that p. You answer the question whether you believe that p by introspecting or reporting whether you are disposed to feel that p when you consider the issue. You answer the question whether you accept that p by making or reporting a decision, or by framing or reporting an intention, about the foundations of your reasonings. Acceptance concerns not what you (passively) feel to be true but what you (actively) premise to be true. Accordingly acceptances are normally explained as outcomes of reasoning in accordance with rules, whereas beliefs are explained as resulting from the operation of relevant causal factors, such as sensory stimuli.

Belief is thus not a disposition to say that things are thus-or-so, but a disposition to feel that they are. Of course, some people are so talkative that they try to tell you every belief that they have, and perhaps every other mental feeling too–every hope, fear, desire, regret, and so on–unless there is some special reason for keeping it to themselves or you manage to extricate yourself from their garrulousness. Other people are reticent to the point of secrecy, unless there is some special reason for disclosure, and they may have just as many beliefs and other feelings. So a disposition to speak and act as if one feels that p is not a necessary condition for its being true that one believes thatp. Nor is it a sufficient condition for this. A person may pretend, in what he says and does, to believe that not-p, while in fact he believes that p.

Belief and acceptance, though different in their nature, have a natural tendency to run together. On the one hand, belief that p is at least a prima facie reason for accepting that p. We often need to presume that our subconscious belief-inducing mechanisms, especially those of

perception and memory, have operated veridically. On the other hand, acceptance that p tends to promote belief that p, as Pascal argued in the case of religious belief.[3] And this natural tendency for states of belief and acceptance to be linked together in everyday experience is probably the main factor in generating a tendency for the two concepts to be confounded with one another in recent philosophical reflection.[4]

But there are a number of contexts in which belief and acceptance do not in fact coincide. Some of these are quite humdrum. For example, even though a person who accepts nothing that he believes is intellectually self-paralyzed, a person who accepts everything that he believes is recklessly uncritical, since there are so many possibilities of accidental error through visual illusion, mishearing, misunderstanding, and so on. Or, perhaps a person can't really believe that his hero is a crook, though the evidence is so strong that he decides to accept it. Also there are often ethical or prudential reasons for accepting that p which are not good reasons for believing that p–as when professional considerations require a defending lawyer to deem, in the context of a trial, that his client is not guilty, though in his heart he believes the client guilty. If the lawyer does not, whenever it is appropriate, adopt the policy of premising that his client is falsely accused, he will be doing less than he can to protect his client's interest.

More importantly, belief, as the disposition to have feelings of a certain kind, is involuntary. Beliefs just come over you or grow on you. You can plant them in others' minds, not in your own. But acceptance, as the adoption of a policy, is a voluntary act. So, in general, people are held responsible and accountable for what they accept, not for what they believe. Of course, people are often at fault for not acquainting themselves with relevant facts or for not acquiring relevant disciplines. But people are not to be blamed if they still lack the belief that p, even after acquainting themselves with evidence and inferential procedures that are normally adequate to produce the belief that p.

Another important point is that, while the explicit or direct act of acceptance involves a conscious and voluntary adoption of a policy about premises, a person may be said to accept implicitly or indirectly all the deductive consequences of each of the propositions that he accepts explicitly and any of their conjunctions, so long as he is himself aware of, or able to work out, the deducibility of those consequences. Beliefs carry no such deductive commitment.

Once the distinction between belief and acceptance has been properly formulated, it can be used to clarify a number of epistemological issues.[5] For example, contemporary analytical philosophers commonly suppose that everyday explanations of human behavior operate within a

framework of the agent's beliefs and desires.[6] But a great deal of human behavior can only be explained by reference to what the agent voluntarily accepts, not what he involuntarily believes, and in terms of the goals or policies that he voluntarily adopts, not the desires which he involuntarily feels. Again, we have to ask whether the progress of scientific inquiry is best furthered by scientists believing their favored theories or accepting them. And it turns out that the latter is more likely to preserve open-mindedness to new ideas. Acceptance in the context of scientific inquiry also integrates better with the conception of theories in the advanced sciences as idealizations rather than descriptions. Moreover, only when we have thoroughly explored the different roles that belief and acceptance are conceived to play in self-deception can we gain a proper understanding of the concept of self-deception.

3. Should Triers Of Fact Declare What They Believe Or What They Accept?

Let us turn now to the particular issue before us–the issue about verdicts. In law courts, questions of fact are sometimes decided by a single judge and sometimes by a panel of judges, sometimes by one or more judges afforced by one or more expert assessors, and sometimes by a specified number of lay jurors. But what do such triers of fact properly give the world to understand when they announce their decisions? Do they imply, for example, in a criminal case that they are disposed to have a feeling of conviction that the defendant is guilty? Or do they imply that they have rightly adopted a policy of premising that this is so? Are they claiming that they have been uncontrollably overwhelmed by a belief? Or are they declaring that they treat the evidence and arguments put before them as adequate reason for accepting that the defendant is provably guilty?

The point is not an unimportant one, because considerations of belief and considerations of acceptance may lead in different directions. For example, a juror, because of some racial prejudice that he is powerless to discard, may experience deep and uncontrollable feelings of conviction that the defendant has been proved guilty. Yet, if he carefully evaluates the strength of the evidence before the court, he may think it reasonable for him not to accept that proposition. Or a juror may firmly believe that the defendant has been proved guilty, but be unwilling to accept that he should convict, because he deems the law under which the defendant is being tried to be unjust. Or he may believe that a key witness for the defense is lying, but be incapable of giving any other reason than this belief for accepting that the witness is lying, so he hesitates to accept

that proposition even though he believes it. On the other hand, a juror may be willing to accept what follows in the instant case from some rule of evidence–such as a presumption of guilty knowledge or a requirement of corroboration–but be quite unable to have the corresponding belief or absence of belief.

Actually there is a threefold issue here.

One kind of question raised is about how different legal systems require their triers of fact to operate. For example, so far as English criminal courts are concerned, was Lord Diplock right in *Walters v The Queen*[7] to approve of Lord Goddard's statement in *Rex v Kritz*[8] that the judge should make the jurors understand that "they have to be satisfied and must not return a verdict against a defendant unless they feel sure"? And was Sir Owen Dixon right in what he said in the Australian case of *Briginshaw v Briginshaw*?[9] He said:

> The truth is that, when the law requires the proof of any fact, the tribunal must feel an actual persuasion of its occurrence or existence before it can be found. It cannot be found as a result of a mere mechanical comparison of probabilities independently of any belief in its reality.[10]

If Diplock, Goddard, and Dixon were right, the English and Australian systems certainly require a convicting jury to believe that the defendant's guilt has been proved beyond reasonable doubt, whether or not it also accepts that proposition. On the other hand the old Carolingian code, with its exact proof-metric, clearly laid down conditions under which the guilt of an accused had to be accepted whether or not the judges believed it, such as when there were adequate grounds for torture and a confession was extracted.[11] But these are questions about existing law and are outside the scope of my present inquiry. They ask: what are, or were, the actual rules of law at a particular time and place?

A second kind of question that arises here concerns the ways in which triers of fact operate in practice, whether or not those patterns of operation conform to the relevant legal requirements. And about such questions there is obviously room for sociological or psychological inquiry, which is outside my present concern.

The third kind of question has a more philosophical orientation. Existing rules and practices may not be ideal, or even consistent. So we can ask also: What does justice, or the public good, require? If such an unqualified question is too simplistic, we can ask instead: What conditions are relevant to the requirements of justice and the public good, with respect to the belief/acceptance issue? Should belief that p be sufficient for a verdict that p? Should it be necessary for such a verdict?

It is easy to see why the present state of English and Australian law, as viewed by Diplock, Goddard, and Dixon, could seem well motivated in relevant respects. In the English and Australian jury systems a group

of people recruited from the populace at large has no other function to perform than the delivery of a verdict on the facts presented to it. Such a lay jury, from the point of view of the general public, operates rather like what psychologists call a black box. It receives an input of data from testimony, behavior of witnesses, advocates' arguments, judicial notice, statements of law, and summation. It delivers an apparently nonrandom output in the form of a verdict. But it discloses nothing then about the intervening process, except whether or not the verdict is unanimous. Neither the jury as a whole nor its individual members pronounce any reasons for the verdict, and individual jurors are discouraged from later describing what happened during their deliberations. Any appeal on a factual issue has therefore to be grounded on defects in the evidence, procedure, or summing-up, rather than on identifiable pieces of mistaken reasoning by the jury. So on the information officially available, all one can say is that, in the circumstances present, what is put into the black box somehow determines its output. There is certainly no official ground for supposing that any reasoning has taken place, let alone that the reasoning has had this or that specific structure (except perhaps insofar as the judge may have been consulted about some point of law or fact that has arisen during the jury's deliberations). Hence one might well think that nothing is gained by setting up legal rules about how a jury should deliberate: such rules would not be enforceable by the court. Any institutional safeguard against the influence of illegitimate considerations–corruption, intimidation, racial or sexist bias, popular gossip, prejudgment by the media, and so on–must lie in the surveillance of individual jurors or in the size and composition of the jury or of its requisite majority. Accordingly, because belief is explicable in terms of causation and acceptance is explicable in terms of reasoning, such a system seems to favor the view that a jury's verdict expresses what the jurors (or a permissible majority of them) believe, rather than what they accept. And within this kind of institutional framework it is quite intelligible why some professional judges might come to hold it just that conviction, or feeling sure, should constitute the requisite warranty for a jury's verdict of guilt. Indeed it may also be much easier for a lay juror to introspect his own feelings than consciously to compute the validity of proofs.

However, the black-box character of the jury's deliberations is not the only factor that may be held to affect the issue, even if we confine our attention to the common-law tradition.

Another important factor here is the tendency of juries to reflect the prevailing values of their community in their decisions. If the penalty that would follow conviction is seen as being too severe, or if the alleged

crime is regarded, in the circumstances, as an honorable deed or a public service, a jury may give a verdict of acquittal despite strong evidence of guilt. What is happening here? You may be tempted to say that in pronouncing its verdict the jury purports to believe that the defendant has not been proved guilty, while it actually believes that he has. But that description of the situation imputes deliberate dishonesty or deception to the jurors. Such an imputation may sometimes be correct. But normally no one is deceived about the jurors' actual state of mind in such a case. So it seems wrong to describe their state of mind in terms that imply dishonesty. Indeed, a juror need not think of himself as being dishonest in this way if the correct description of his conduct is to say that he deems the defendant not to have been proved guilty and accordingly acquits him, while nevertheless believing him to have been proved guilty. The jury would then be accepting, on ethical or pragmatic grounds rather than on cognitive ones, that the accused is not guilty, while believing perhaps quite firmly in his or her guilt. So the verdict would declare what the jury accepts, not what it believes. Of course in many, if not most, jurisdictions this kind of behavior conflicts with a jury's formal obligations. But the fact that–in suitable circumstances–a jury is likely to behave like this is one reason for the continued popularity of the jury system.[12] Correspondingly, that popularity tends not to sit well with the Diplock-Goddard-Dixon type of view.

But even the existence of an official ban on such morality-based verdicts does not necessarily favor the view that juries should declare what they believe rather than what they accept. Perhaps the standard procedure for banning those verdicts is to require the jury to swear to heed the evidence. Etymologically, after all, a juror is an oath-taker. Now suppose a juror does have to take an oath, like the English one: "I will faithfully try the defendant and give a true verdict according to the evidence." What has to be kept in mind is that swearing an oath to perform a particular act commits one to believing, at the time the oath is taken, that it is within one's power to perform that act. For it follows from the existence of this commitment that what an English juror swears to do is to *accept* the consequences of the evidence. He cannot swear that his *beliefs* will be in accordance with the evidence because his beliefs are inherently involuntary and not within his power to create. Thus, where jurors have to take an oath, the argument from the black-box character of their deliberations is overridden by the implications of their oath-taking. Even though the former argument gives some support to the view that the jury's *belief* should be what is crucial, the requirement of an oath implies that what ought to be crucial is what the jury *accepts*.

Of course, though a juror cannot control his beliefs, he can in principle control certain other factors (such as his attention to witnesses' testimony or to the judge's admonitions) that may or may not cause beliefs about the adjudicated issue to arise in him. Indeed, by controlling those other factors he may even increase the probability that he has true beliefs in regard to what has been proved about the defendant. But vowing to do what one can to produce the right beliefs in oneself is not the same as vowing actually to have the right state of mind. Compare how my promising to try to arrive by 5:00 p.m. is not the same as my promising *tout court* to arrive by 5:00 p.m. The latter commits me definitely to accepting that I can arrive by 5:00 p.m., while the former does not. Correspondingly the juror's oath is to give a true verdict according to the evidence, not just to try to give a true verdict according to the evidence.

Accordingly, if the interest of justice or of the public good is promoted by a requirement that triers of fact be sworn, then the nature of the oath that would be appropriate for jurors entitles acceptance, rather than belief, to be considered the mental state that–in the interest of justice or of the public good–should underlie jury verdicts. Perhaps someone will object that what a juror should swear to do is to pronounce in court (or in the jury room) the verdict that he actually believes in his heart to be correct. The requirement of an oath, on this view, should be aimed just at preventing any discrepancy between what the trier of fact believes and what he says. A juror's oath, like a witness's oath, should promote candor and honesty.

But the objection will not stand. And my argument against the objection is not that existing oaths, like the English one, are not of the kind said by the objector to be needed (although they indeed are sometimes not of this kind). Rather my argument is that oaths which are of this kind do not cover all the undertakings that are needed. What good is it for a juror to tell the court exactly what he believes if his belief is the result of corruption, intimidation, bias, prejudice, inattention, or fallacy? How can justice be done unless the conclusion accepted by a juror is adequately defensible? Whatever a juror accepts should be accepted for the reasons supplied by the facts before the court. So the oath taken by a juror must be directed not only towards preserving the conformity of what he says to what he thinks, but also towards preserving the conformity of what he thinks to what the facts before the court are. And for the latter purpose–just as well as for the former–it must be a juror's voluntary acts of reasoned acceptance that count, not his involuntary states of instinctive belief.

Another important factor is the extent to which the legal system regulates proof. Any such regulation tends to attract attention to the possibility of a divergence between belief and acceptance as the mental foundation for a juror's verdict. This is because, when a rule is stated or applied by a presiding judge, it tends to focus a juror's attention on what he is, or is not, entitled to accept in virtue of it irrespective of what he believes. For example, where a proof depends at any point on a presumption, such as a presumption of legitimacy or of guilty knowledge, a lay trier of fact may be able to accommodate the presumed truth as a reason for accepting the proposed conclusion but be quite incapable of coming to believe that conclusion. Again, where the legal system requires corroboration for a certain type of testimony, such as the testimony of young children, a lay trier of fact may be inclined to believe the conclusions that flow from some piece of uncorroborated testimony but nevertheless be willing not to accept those conclusions as a basis for his verdict. This would be even more likely to happen where the relevant legal rule seemed rather a technical or artificial one, as with the old English rule that testimony requiring corroboration cannot corroborate testimony requiring corroboration.[13] Similarly, if in the same trial some piece of evidence is admissible on one issue but not on another, it may be relatively easy for jurors to bear this in mind in relation to the conclusions that they accept, even though their beliefs are inevitably influenced by it on both issues. Indeed it is plausible to suppose that exclusionary rules, which help to determine acceptance-worthiness, have sometimes entered the law of evidence in particular legal systems just because of the dangers that would exist if verdicts by lay tribunals were to be based on instinctive belief rather than considered acceptance in cases where the tribunal has unfortunately heard some forbidden evidence and is told to disregard it.

It follows that a legal system with fairly extensive exclusionary rules of evidence does not fit in easily with the view that verdicts should express belief rather than acceptance. Correspondingly, where the nature of the judicial process is such as to generate a need for rules of evidence, one would expect triers of fact to declare what they accept rather than what they believe. So, if an adversarial–or predominantly adversarial–system tends to require more rules of evidence than an inquisitorial – or predominantly inquisitorial-system, the adversarial system may appropriately have a stronger tendency to subordinate belief to acceptance.

But it does not follow that an inquisitorial system may not also work towards acceptance rather than belief. Such a system commonly (though not universally) operates without a separate jury, with a single judge as both trier of fact and interpreter of law. Conceivably the system

might allow or require the judge not to weigh up the available facts or to try to derive a reasoned conclusion. The inquisitorial judge's duty might be merely to consult his convictions after he has ingested all the relevant data. He would then be treating his subconscious mind as a kind of black box that takes evidence and legal argument as input and delivers nonrandom verdicts as output. But an inquisitorial system does not have to operate like this. Such a system could certainly expect to inspire greater public confidence in the correctness of its verdicts if its verdicts had to be supported by reasoning and the reasoning also had to be published. Where the trier of fact does not have to argue out the issues with colleagues, he ought at least to argue them out with himself, as a precaution against obvious risks of bias, prejudice, overconfidence, or accidental paralogism.

Of course, it is in principle always possible for a solitary trier of fact to consult his private convictions first and construct his publicly announced justification afterwards, so as to present the appearance of having weighed up the arguments before coming to a conclusion. And this procedure, where it exists, may still be construed in terms of the domination of acceptance over belief, since it may be taken to reflect a determination on the judge's part to accept the conclusion that he believes instinctively to be correct rather than the conclusion at which he might arrive through conscious reasoning. But in such a process acceptance does not play as crucial a role, because it rides in only on the back of belief.

In sum, then, legal process does best to aim at verdicts that embody appropriately reasoned acceptance. In this the legal process resembles scientific inquiry which, I have argued elsewhere, does best to aim at the acceptance of explanatory and predictive theories.[14] In both cases the knowledge pursued requires adoption of an active, rather than passive, attitude towards resolution of the issues. And just as in scientific inquiry surrender to belief may harm the progress of science by blocking reconsideration of issues that need to be reconsidered, so too in trials of fact, surrender to belief may obstruct the pursuit of justice by making it easier for illegitimate considerations to influence verdicts. If justice is to be done, the canonical purpose of advocacy ought to be to provide proofs or disproofs that a rational trier of fact could accept. The canonical purpose ought not to be to engineer a desired state for the trier's credal feelings, even though this may in practice result. More specifically, in the context of Anglo-American law, any application of either of the two familiar standards of proof ought to be subject to acceptance rather than belief. For a criminal conviction the jury needs to accept that guilt has been proved beyond reasonable doubt, and in a civil court

the plaintiff should win if and only if the trier of fact accepts that the plaintiff has the balance of evidence in his favor.

Of course, belief does nevertheless have one important role within the structure of courtroom thinking. In the standard case, belief is the appropriate attitude towards the data. If a witness testifies that such-and-such is a fact, he implies possession of a belief that what he testifies is a fact. Also the trier of fact must believe whatever propositions he takes as ultimate premises–such as propositions about what the witnesses say or about how they behave under cross–examination. Here again legal process is like scientific inquiry, where those who produce the original reports of experimental or observational data are assumed to believe that those reports are true. However, just as a scientific textbook may report facts that the author of the textbook has not himself observed but nevertheless accepts as genuine, so too an expert witness, as distinct from an eyewitness, reports in court what he and his colleagues accept within their relevant field of expertise.

4. If different standards of proof are formulated in terms of different levels of probability, are those standards to be understood as expressing different strengths of belief?

There is an important objection that might be made here. Legal systems often operate with more than one standard of proof, it might be said, the attitude of mind that underlies a fact-finder's verdict must be one that admits of gradation. In Anglo-American law a higher degree of that attitude is requisite for a verdict against the defendant in criminal cases than in civil ones, and in civil cases a winning plaintiff must sustain a higher degree of it than the defendant. Now acceptance, runs the objection, does not come in degrees. You either adopt the policy of premising that p or you do not. But beliefs can be weaker or stronger, in accordance with the persuasiveness of the evidence. Moreover, degree of belief, if appropriately coherent, has the mathematical structure of a probability, as Frank Plumpton Ramsey, Bruno de Finetti, and others have shown.[15] So, if we suppose that what underlies any honest verdict is a certain degree of belief, we can then interpret the requisite standard of proof as the level of mathematical probability which this degree of belief ought to constitute in the type of case in question. We have therefore a good reason to prefer the type of system to which, on the Goddard-Diplock-Dixon view, English law belongs. Verdicts are best based on corresponding beliefs, not on corresponding acceptances.

In order to answer this objection we need to consider carefully what state of mind is at stake in a subjectivist analysis of probability. According to such an analysis a spoken or written assessment of probability is to be understood as describing or avowing some cognitive feature of the assessor's own state of mind. But we need to ask: Is that feature to be identified with a passive or with an active attitude of cognition? Certainly Ramsey explicitly identified a person's assessment of the probability that p with a declaration of his or her degree of belief that p[16] (though Ramsey did not think that degree of belief is measurable by intensity of feeling).[17] De Finetti has held a similar view.[18] And in taking this line Ramsey and de Finetti have been followed by most subjectivists.[19] But is such a view just a Humean, positivist prejudice–a historically accidental accretion–which a sophisticated neo-Cartesian subjectivism could discard? Or is reference to belief a necessary feature of any attempt to analyze assessments of probability as declarations about features of the assessor's own state of mind? This issue is particularly important in the present context because those who analyze standards of juridical proof as thresholds of mathematical probability are often inclined to favor a subjectivist, Ramsey-type account of mathematical probability.[20]

It is clearly material at this point to consider first how betting behavior is best explained, since most subjectivists suppose people to reveal their assessments of the probability that p by the lowest odds on which they will bet that p. In fact according to Ramsey, "a person's actions are completely determined by his desires and opinions," where opinions is a synonym for beliefs.[21] So according to Ramsey the action of placing a bet at specific odds is to be explained in terms of the bettor's desire for the goods that are at stake in the bet and the strength of the bettor's belief about the outcome. Indeed Ramsey thinks that strength of belief that p is hardly to be known otherwise than via the believer's lowest odds for betting that p.[22]

But, as I noted earlier, there is also quite a different way of explaining a person's behavior, whether the behavior be concerned with bets or with anything else. Specifically, some of a person's behavior may have a rational explanation–an explanation in terms of the premises that he accepts and the goals that he adopts. That type of acceptance-based rationale has several advantages over Ramsey's method, for explaining betting behavior. One advantage of the acceptance-based rationale is that it allows for the possibility of insincerity. This is important because a person's betting behavior can be designed to conceal his real belief about the outcome rather than to reveal it.

A second point that has relevant implications for subjectivism is the importance, already remarked, of accepting a scientific theory rather

than believing it.[23] Subjectivists are normally Bayesians in their methodology for science and other intellectual inquiry, measuring evidential confirmation for a theory by the extent to which the posterior probability of the theory (that is, its probability on the evidence) exceeds the theory's prior probability (that is, its probability before the evidence is taken into account).[24] That posterior probability is then treated as a highly relevant consideration in determining whether or not the theory should be accepted. But, if a scientist who accepts the theory that p does best to lack any belief that p, as I suggested earlier, then his personal evaluation of the posterior probability that p should not amount to a declaration of his degree of belief that p. A scientist's declaration of the acceptance-worthiness of a theory should not depend on the strength of his belief that the theory is true. In sum, the view that in science and other intellectual inquiry acceptance rather than belief is the right attitude to have towards a preferred theory, carries with it the view that the probability of the theory should not be construed in terms of degree of belief.

A third way in which the distinction between belief and acceptance affects the subjectivist conception of probability is that it engages quite sharply with the issue of coherence. Since wagers are voluntary undertakings, which do not necessarily reflect the strength or weakness of the bettor's underlying mental feelings, the bettor's offer of his lowest odds on the proposition that p claims to reflect what he accepts in his mind as being appropriate about the terms on which he will bet rather than to reflect the strength or weakness of his underlying mental feelings. And it would not be possible to set up a rational methodology for assessing probabilities on the basis of such subjectively assigned betting odds unless the requisite coherence of those assignments–that is, their resistance to a Dutch book (in which the bettor loses money overall whatever the outcomes were conceived in terms of coherent acts of acceptance rather than in terms of coherent states of belief. Coherence, in this context, has to be an attainable norm. Otherwise the process involving it could hardly be the concern of methodology. And if coherence is to be an attainable norm, the attitudes required to attain it must be within a person's voluntary power, as beliefs are not. So, whatever Bayesians *say* about strength of belief, what their methodology must actually operate with is acceptance. The strength of a person's various interrelated beliefs may in fact be distributed quite rationally. But, if so, he may just be lucky. What Bayesian methodologists can recommend him to ensure–as a practicable policy for action–is that his *accepted* betting odds cohere when he originates or revises them.

Of course, certain beliefs are indeed relevant here. On a subjectivist view one legitimate reason for a person's accepting such-or-such odds for betting that p–if he or she can do so coherently–may be because he has such-or-such a strength of belief that p. The subjectivist thesis is, then, that in these circumstances the belief may justify the acceptance: the belief's occurrence is a cognitively relevant consideration. But to say this is not to imply that the strength of a person's belief that p actually constitutes his evaluation of the probability that p. What justifies a particular cognitive attitude cannot be identical with what constitutes it (any more than the falling barometer that justifies a storm alert is identical with the alert). That becomes very clear where what justifies your acceptance of such-or-such odds on its being true that p is not the strength of your belief that p but some cognitively irrelevant factor, such as your aim to attract others to bet with you on the issue.

A fourth point here concerns the monitoring of biases. If subjective probabilities are held to be constituted by acceptance, not belief, people are encouraged to take a more active view of how these probabilities should be evaluated. No doubt one piece of evidence that a person would take into account in any such evaluation, as remarked in the previous paragraph, is the strength of his own passive feeling of confidence in the relevant outcome. Treating his belief system–that is, the causal system generating his belief–as a black box, a person is entitled to record the content and strength of his belief system's output when available information is fed into it. A person may indeed form a provisional assessment of appropriate odds on this basis. But if he is prudent he will adjust this provisional assessment in the light of what he believes about his own biases, such as tendencies towards overconfidence or underconfidence in certain kinds of outcome. That is to say, he might take those beliefs as reasons for accepting in the end an assessment of subjective probability that differs somewhat from the assessment that the black-box procedure on its own would generate. So by making acceptance ultimately crucial, rather than belief, we allow room for a person to exercise conscious self-control, even prior to ensuring coherence. And because acceptance is voluntary, a person can thus also be considered responsible for inadequacies of self-correction in his fundamental assessments of probability. Indeed, so long as subjective probability is supposed to be rooted in an involuntary degree of belief, a subjectivist analysis is unsuitable for any situations in which a person, like a weather forecaster or a market analyst, may be held culpable for his or her erroneous assessments of probability. But people can be held accountable for probability assessments made in subjectivist terms, if, instead, what is ultimately crucial is what they accept. The distinction between belief and acceptance pays

yet another dividend in this context. It shows what is wrong with Bas C. van Fraassen's argument for the implausible conclusion that belief is voluntary.[25]

Van Fraassen begins by arguing that someone operating over time with a subjectivist, Ramsey-type conception of probability[26] could only avoid a Dutch book being made against him on either of two conditions. One of the two possible conditions is that the person concerned has no beliefs about the reliability of his own future judgements. The other condition is that the person has a degree of present belief, n, about a certain future outcome on the supposition that he will in the future have a degree of belief, m, about that outcome, if and only if $n = m$; that is, for any degrees of probability n and m, the present probability of the outcome–given that the person will in the future assign it a probability m–is n, if and only if $n = m$. The former of these two conditions would normally be too restrictive: it confines the ability to avoid a Dutch book to implausibly few people. So, van Fraassen argues, the second condition has to be satisfied. And the only way to reconcile oneself to this is to see the latter condition as a form of commitment to stand behind one's own commitments. Consequently, since commitments have to be undertaken voluntarily, we need to look on any commitment-carrying gradations of probability–when probability is construed in subjectivist terms–as also being voluntary. And if such gradations are gradations of belief, concludes van Fraassen, then belief itself must be voluntary.

However, van Fraassen's argument clearly collapses if the subjectively construed assessment of the probability that p has in any case to be treated as a gradation of the lowest accepted odds that p. Van Fraassen's two conditions for avoiding a Dutch book–in a sequence of bets, over a period of time–must then be reformulated in terms of acceptance, which is admittedly voluntary, and no conclusions can be drawn from either of van Fraassen's conditions about the nature of belief. Indeed, when appropriately reformulated, the argument must rather be taken to reinforce the thesis that any subjectivist measure of probability must be related to acceptance, not belief. That is to say, if belief is indeed involuntary, as I have been assuming, van Fraassen's reasoning may be viewed as a *reductio ad absurdum* proof for the thesis that the degree of a person's subjective probability that p should be identified not with his degree of belief that p, but with his accepted assessment of the lowest odds for betting that p.

My reasoning here is not directed towards defending subjectivism against the arguments of those who advocate objectivist accounts of probability-relative frequency analyses, propensity analyses, and so on. Nor is my reasoning directed towards championing the pluralist account

that I myself prefer (an account that not only admits a certain variety of interpretations for the standard probability function but also admits a certain variety of nonstandard probability functions).[27] Instead my point is just that *if* the subjectivist option is to be stated at all–whether in relation to jury verdicts or other issues–its statement should discard any reference to degree of belief and take acceptance to be the mental attitude that is crucially involved. In this way we provide an appropriate explanation for a person's betting behavior; we fit in with the cognitive attitude that it is appropriate to adopt towards a preferred scientific theory; we leave room for methodological promptings and attributions of responsibility that are aimed at the pursuit of coherence and the avoidance of bias; and we can escape a diachronic Dutch book without committing ourselves to the paradoxical doctrine that belief is voluntary.

An important point to notice, however, is that we do not then just replace the term *belief* in the conventional subjectivist analysis by the term *acceptance*. The issue is not a merely verbal one. Though acceptance, like belief, may vary with the occasion, acceptance, unlike belief, may not vary in degree. A person may take the proposition that p as a premise for business purposes but not in his private life, just as his belief that p may come and go with the presence and absence of some persuasive informant. But acceptance does not admit of any kind of internal variation other than in its content. Of course, acceptance-worthiness is a matter of degree. But it is not a subjective state of mind, like degree of belief. Also, where 'p' and 'q' stand for any two propositions, a person's inclination to take the proposition that p as his premise may be stronger than his inclination so to take the proposition that q. But inclinations are prior to decisions or the formation of intentions, while acceptance that p executes a decision or intention. Acceptance that p cannot itself be stronger than acceptance that q in the way that one's inclination to accept that p can be stronger than one's inclination to accept that q. To accept that p is to adopt the policy of taking the proposition that p as a premise in appropriate circumstances, and you either accept that policy or you don't. That is why philosophers of science have argued about the level of a theory's probability, confirmation, support, and so on, that may be appropriate for its acceptance. Probability, confirmation, support, and so on, are matters of degree in terms of which, alongside other considerations, scientists may be able to fix a threshold of acceptance for the theory. In contrast, it is because beliefs are dispositions to have certain feelings that they can vary in strength with the intensity of those feelings. Indeed the variation in possible intensity of feeling of belief, from mere suspicion to overwhelming conviction, is obviously comparable with such other variations of mental feeling as that from

mere irritation to overwhelming fury or that from gentle satisfaction to fierce exultation.

Consequently, when the probability that p is measured by the lowest odds at which a person is prepared to bet that p, we have to think of the probability that is thus measured as an accepted gradation of something that the person accepts about the proposition that p rather than as a gradation of his or her acceptance that p. You can perhaps first accept that it is appropriate for you to bet only on rather long odds that p, and later accept that it is appropriate for you to bet on shorter odds that p. But what thus changes in degree is the length of the odds that you accept, not your own act of acceptance.

Moreover, this attribution of change to the object of gradation – i.e. to what is graded – has important consequences for the complementationality of the mathematical function involved. According to the laws of the mathematical calculus of chance, if the probability that p is very nearly minimal, the probability that not p has to be very nearly maximal.[28] But, if a person's belief that p is so weak as to be almost nonexistent, it certainly does not follow necessarily that he has a very strong belief that not-p. There might just be very little evidence to sway him either way. So the degree-of-belief account of subjective probability tends to clash with the complementationality principle for the probability of a negation. So too does any account in terms of inclination or willingness to accept, or in terms of the degree of wholeheartedness with which one accepts. A person may be very little inclined to accept that p as well as being very little inclined to accept that not-p: perhaps he just wants to keep an open mind on the issue. An account in terms of acceptable odds, however, does not face any difficulty here. To accept that it is appropriate for you to bet on long odds that p is necessarily equivalent to accepting that it is appropriate for you to bet on short odds that not-p.

It follows that even when you make due allowance for the existence of different standards of proof and then choose (as I would not)[29] to formulate these gradations of proof in terms of a subjectively construed probability function, there is still good reason to conclude that what should underlie a fact-finder's verdict is an attitude of acceptance rather than of belief. In short, subjective probability judgements should not in any case be interpreted as expressing degrees of belief; a fortiori, such judgements should not be interpreted in this way when used to represent a certain level of proof.

5. Conclusion

The epistemologically important distinction between passive and active cognition needs to be kept carefully in mind when we consider what it is that a juror's verdict should ideally declare. The Goddard-Diplock-Dixon type of view may or may not be correct as a proposition about the current state of the law in England or Australia. But it does not seem to be correct as a coherent and defensible account of what a system of law ideally requires from its triers of fact. For that purpose it is necessary to develop an account in terms of acceptance, not belief. If some critics still prefer an account in the latter terms, I hope at least to have sharpened up the issue so that they can become aware of the main counterarguments that they must meet.[30]

Notes

1. See R. Descartes, *The Philosophical Works* 144-57 (E. Haldane & G. Ross trans. 1911).

2. D. Hume, *A Treatise Of Human Nature* 624 (L. Selby-Bigge ed. 1888).

3. B. Pascal, *Pensées* 95 (J. Warrington trans. 1960).

4. See, e.g., A.I. Goldman, *Epistemology And Cognition* 324 (1986); Maher, "The Irrelevance of Belief to Rational Action", 24 *Erkenntnis* 363 (1986).

5. A longer but as yet partial discussion of the belief/acceptance issue is to be found in Cohen, "Belief and Acceptance", 98 *Mind* 367 (1989). A fuller treatment will appear in L.J. Cohen, *An Essay On Belief And Acceptance* (forthcoming 1992).

6. See, e.g., D. Davidson, "Actions, Reasons, and Causes", in *Essays on Actions And Events* 3, 3-4 (1982).

7. [1969] 2 App. Cas. 26, 30 (P.C. 1968) (Jamaica).

8. [1950] 1 K.B. 82, 89 (1949).

9. 60 C.L.R. 336 (Austl. 1938).

10. *Id.* at 361.

11. See J. Langbein, *Prosecuting Crime In The Renaissance: England, Germany, France* 188 (1974).

12. See A.A.S. Zuckerman, *The Principles Of Criminal Evidence* 32-4O 1989)

13. See Director of Pub. Prosecutions v. Hester, [1972] 3 All E.R. 1056; *see also* A.A.S. Zuckerman, *supra* note 12, at 44.

14. Cohen, *supra* note 5, at 385-86.

15. See F.P. Ramsey, "Truth and Probability", In *The Foundations Of Mathematics And Other Logical Essays* 156, 174 (1931); de Finetti, "Foresight: Its Logical Laws, Its Subjective Sources", In *Studies In Subjective Probability* 93 (H. Kyburg & H. Smokler ed. 1964).

16. F.P. Ramsey, *supra* note 15, at 174

17. *Id.* at 169.

18. de Finetti, *supra* note 15.

19. See, e.g., C. Howson & P. Urbach, *Scientific Reasoning: The Bayesian Approach* 57 (1989).

20. See, e.g., "Kaye, The Laws of Probability and the Law of the Land", 47 *U. CHI. L. REV.* 34, 44 (1979).

21. F.P. Ramsey, *supra* note 15, at 173.

22. *Id.* at 183.

23. See *supra* text accompanying note 14.

24. See, e.g., M. Hesse, *The Structure Of Scientific Inference* 103-50 (1974)

25. See van Fraassen, "Belief and the Will", 81 *J. Phil* 235, 235-36 (1984).

26. See F.P. Ramsey, *supra* note 15.

27. *See* L.J. Cohen, *An Introduction To The Philosophy Of Induction And Probability* (1989).

28. *Id.* at 13-17.

29. *See* L.I. Cohen, *The Probable And The Provable* 49-120 (1977).

30. I am grateful to Dr. A.A.S. Zuckerman for some very helpful comments on an earlier draft of this paper.

20

ARE THERE ETHICAL REASONS FOR BEING, OR NOT BEING, A SCIENTIFIC REALIST?

I

There are a variety of different kinds of normative questions that may be asked about science. We can ask, for example, whether laws should be conceived as stating no more than constant conjunctions? Or how should experimental controls be structured? Or is evidential support measurable? Or do theories function only as anti-realists claim that they do – viz. as instruments for the prediction of observable events? Or is there only one correct pattern of scientific finding, or, as some readers of Alasdair Crombie's *Styles of Scientific Thinking in the European Tradition* might suppose, are there several? These and their like are one familiar group of normative questions about one aspect of scientific enquiry. Let us call this the "epistemological" aspect, because it is this kind of question that needs to be answered in order to make progress towards resolving the larger issue that may be conveniently summarised in the question "What is scientific knowledge?"

A second group of normative questions may be said to concern the "ethical" aspect of scientific enquiry rather than the epistemological. What good can scientific discoveries achieve? What moral constraints should control the choice of options in experimental design? Ought the choice of research topic to be affected by the good, or bad, consequences that success in the research might have? Is there only one system of ethics with which science is compatible, or are there several?

Other groups of normative questions about scientific enquiry also exist. For example, questions may be asked about the social structure that is optimal for scientific progress – questions influentially raised by Francis Bacon in his *New Atlantis* and still exercising the minds of administrative and grant-giving organisations. Or questions may be asked about conventions in writing scientific papers, such as conventions about how names are ordered in the title of a jointly written paper. And there are also, corresponding to each of the above groups of normative questions, important groups of factual questions about science, which are not always easily separable, as in the history of scientific discovery, the history of scientists' value-systems, the history of conventions in scientific literature, and so on.

I am concerned here, however, only with the first two groups of normative questions and, in particular, with a problem about the relationship between those two groups of questions. I want to address the problem whether there are any connections between questions of the two kinds, such that some reasonable answers to a question of the second kind (the ethical kind) can support some particular answer to a question of the first kind (the epistemological kind). And I shall concentrate on only one possible question of such a type, viz. are there ethical reasons for being, or for not being, a scientific realist? Of course, some people may have an immediate intuition that such reasons cannot exist. But by them my purpose can be regarded as an attempt to explore the rational basis for such intuitions.

II

Let us consider first an argument for answering this question in the negative. The argument begins with the claim that, if some moral judgement is a member of a set of jointly sufficient and severally necessary premisses for the support of scientific realism, then any assertion of scientific realism must have a moral element. But – the argument continues – all moral judgements in science have a certain structural character: specifically, they all apply some general moral principle to the particular circumstances of science. And scientific realism does not have this structure: it makes no such application. Therefore, concludes the argument, scientific realism has no moral content and there are no ethical reasons for being a scientific realist.

Now here there seem to be two premisses that need substantiation. One is the premiss that all moral judgements in science have the specified structural character: the other that scientific realism does not have this character. Let us consider these two premisses in that order.

There are certainly no immediately obvious exceptions to the thesis that all moral objections in science involve the application of some quite general moral principle to the particular circumstances of science. Consider, for example, how policies for scientific research are standardly evaluated. What normally counts is their probable nett contribution to the achievement of human needs, wishes and benefits, even when this involves nothing but the satisfaction of curiosity. The contribution may vary in its degree of immediacy. "Experiments of light", as Francis Bacon called them, may themselves be justified by the leads they supply to "experiments of fruit" . Admittedly animists, animal lovers, or Gaia worshippers, may attach rather less importance to human welfare and evaluate other ends more highly. But, whatever the nature of the values that scientific activity seeks to promote, their relevance is not peculiar to science.

An analogous situation arises when the experiments needed to test a particular hypothesis cannot be carried out without causing what are thought to be morally unacceptable side-effects, such as through vivisectional experimentation in animal physiology or when misleading information is given to human subjects in certain kinds of psychological experiments. Scientific enquiry runs the risk of ethical involvement here because the processes, structures, etc. that are experimented upon are associated with organisms that may suffer harm to their interests thereby. But there are many other ways – outside the ongoing process of scientific enquiry – in which such interests may suffer similar harm: it is not only in experiments that animals get hurt or people get deceived. So the ethical downside here is due to the situation's being an instance of a general category of situation; and there is nothing peculiar to scientific experiment that is responsible for this downside.

Another such type of situation arises from discreditable behaviour in the handling of information somewhere in the process whereby research findings are established and enter into the canonical literature. A person may make deliberately false statements in applying for a research grant; he may deceive a colleague in a research-team; he may plagiarise from another team or person, or misreport data to a journal, etc. etc. Now the room that exists in science for dishonesty of this kind is due to the fact that, at least since the 17th century and particularly under the inspiration of Bacon's *New Atlantis*, science has normally been thought of as a social activity with consequential rules, presumptions and expectations. Scientific progress is predicated typically of a community rather than of an individual: an individual's achievement is to contribute to this progress. But science is far from being the only type of rule-bound social activity, and correspondingly dishonesty in science owes the source

of its immorality to a principle that governs all inter-personal dealings: namely, the principle that in normal circumstances one shouldn't tell lies to people or mislead them in any way.

Yet another such type of situation arises when a particular line of research is thought to be too dangerous to pursue because of the risk that its findings will inevitably be exploited and generate bad consequences. For example, if the technology of chemical weaponry is researched, some governments can perhaps not be trusted not to use this technology. Or, if the technology of the internal combustion engine has been improved, it may be that some governments cannot be trusted to maintain support for other measures designed to reduce global warming. And here too the ethically relevant features of the situation to be avoided – the unnecessary risk of death etc. – are not peculiar to scientific research. The core of the matter is that one shouldn't facilitate others' misdeeds.

In sum, if we argue inductively from familiar types of case like the four that have just been considered, it looks as though any moral judgement in science is an application of some general moral principle to derive a prescription or proscription in particular circumstances. Science, we may say is ethically heteronomous.

But now let us turn to the other premiss that is needed by the argument under consideration. Does scientific realism indeed have the logical structure that (as the above examples seem to show) is standard for moral evaluation of features of scientific activity? That is to say, does it rest on the application of some general moral principle to the issue about realism in the analysis of scientific theory? And the answer to this question seems to be a negative one For, if we try to identify some general principle in terms of which scientific realism appears as a special case, or derivable consequence, that principle would presumably have to embody some more general form of realism. And any such principle, in vouching for the factuality of what is held to be the case, must inevitably contrast with any kind of moral principle that endorses some range of values or obligations and thereby endorses what ought to be the case rather than what is. The factual is, as it were, a paradigm case of the non-normative.

So, if I am right about the two premisses that I have been considering, it looks as though those premisses provide the necessary foundation for an argument that there cannot be any ethical reasons in favour of being a scientific realist.

III

The situation might be quite different, however, if scientific activity, or some features of it, were independently valuable. For, if scientific

enquiry, or some features of it, were good, or bad, in themselves, and not just as a consequence of a more general moral principle or principles, then the negative argument that I have been sketching would not get off the ground. Since it would not be necessary to endorse some general principle from which a value-judgement about certain scientific activity derived, the lack of such a general principle would not matter, and the possibility of an ethical reason for being a scientific realist would remain. That is to say, perhaps the inductive argument developed in the previous section of this paper is invalid and there is indeed some degree or kind of ethical autonomy in science. Perhaps in appropriate circumstances, and within appropriate limits, and under an appropriate interpretation (such as a realist one might be) we can treat the process of scientific enquiry as an end in itself.

Let us briefly turn aside here to consider a possible analogy with art, since there is an old controversy in philosophy about whether art admits of being good in itself or only of being a means to good results. Thus a medieval altar-piece may have been painted, or a motet may have been composed, so as to facilitate the performance of a religious ritual. Or a statue may have been sculpted, or a poem written, so as to glorify a programme for social reform. In each case the altar-piece, motet, statue or poem may achieve its end just because it has the required features – the icon, tune, size, metre, etc. – that is needed in order to accomplish the religious or political purpose involved. But, though such examples may be cited to illustrate a heteronomous conception of ethical value, their existence does not suffice to establish that this kind of conception is the only admissible one. It makes perfectly good sense to claim at the same time that certain other paintings, songs, sculptures, poems, etc., have an intrinsic value because they conform so satisfactorily to accepted criteria of excellence within their several genres. Indeed three obviously different positions are possible here for anyone who thinks that aesthetic value exists at all. We can hold either that aesthetic merit is always derivable from service to some other purpose, or that it is always derivable from conformity to its own intrinsic criteria, or that sometimes the one holds good and sometimes the other.

Now, in any particular field of research, knowledge and the pursuit of knowledge may be treated analogously to painting, sculpture, music and literature, and their pursuit, within any particular aesthetic framework. Either this value always derives from their utility in the service of some otherwise justifiable purpose. Or their value always derives from their conformity to appropriate criteria intrinsic to science. Or their value derives sometimes from the one and sometimes from the other. Thus instrumentalist analyses of scientific theory may be construed as

assuming that the value of the theory lies in the uses to which it may be put in the construction of inferences, predictions, retrodictions and other intellectual foundations for technology. But, according to Plato, because knowledge of the Forms culminates in knowledge of the Good, there is an ultimate state of knowledge that has to be good for its own sake: knowledge of the Good is not to be sought for the sake of realising some other value. Similarly a Spinozist conception of knowledge – so far as it concerns what it regards as the most perfect form of knowledge – takes the intellectual love of knowledge to be part of the infinite love with which God loves himself and thus takes scientific knowledge to be good in itself. Or – in more modern terms – we might suppose that it is just this that would be done by a religion which worshipped science.

So it at least makes sense to think of scientific enquiry as an activity that can, or even should, be sometimes pursued for its own sake and has an autonomous value. Moreover, when it *should* be so pursued a realist analysis of scientific theory seems more appropriate than an instrumentalist one. For an instrumentalist conception of scientific theory implies the pursuit of science for other purposes than the achievement of knowledge for its own sake. And, conversely, it would then follow that the autonomous ethics of scientific enquiry favours realism rather than instrumentalism in the analysis of fundamental theory.

IV

We need to ask, therefore, how strong are the arguments for saying that science – i.e. the systematic pursuit of knowledge – has, or has not, intrinsic value. One plausible argument for a negative conclusion here might proceed from the premiss that, if a particular line of scientific research cannot reasonably be expected to help to do anybody else any good – i.e. cannot reasonably be expected to help to supply food, cure illness, or satisfy some other important purpose – then the only recognizable motive for pursuing it must be that of selfish curiosity or personal fame. And to use the requisite resources of skilled personnel and expensive equipment in pursuit of such a selfish goal is to reduce the quantity of resources and personnel available for research that might actually do somebody else some good. So, though it may sound alright to treat science, like art, as having an intrinsic value, in fact this is equivalent to pampering the supporters of an immoral policy.

However, the above argument is weaker than it may appear. It assumes that, if science *ought* only to be pursued for the benefits that it may convey to suffering humanity, then this is the motive that will actually determine individual or collective decisions of science policy. But, if there are many potential scientists who are not so charitably

disposed, it is possible that recruitment into scientific research might be poorer in quality, and weaker in quantity, if scientific research were not attributed intrinsic value. Perhaps our communities need a complex system of values in which some activities of a philanthropic character may only get sufficiently promoted because they are encouraged by the value attributed to activities of a non-philanthropic character. Or, to put it more sharply, unless we reject instrumentalism we may fail to attract some of the best scientists into our laboratories.

But does the argument run the other way also? Are there some potential scientists who will work in the laboratories only if they can believe that their work may solve practical problems? Surely yes. So it looks as though the analysis of appropriate scientific motivation will not support a one-sided position on the question about realism and instrumentalism.

V

Let us turn therefore from motivation to methodology. Let us just assume that the scientific enquiries which get funded, and which competent people enthuse about and are honoured for undertaking, are worthwhile irrespective of whether they are pursued for their own sake or for their technologically applicable consequences. Then presumably any methods or practices or philosophical analyses that promote these enquiries are superior to any that do not.

For example, it is often necessary to draw comparisons between the degree of evidential support for one proposed solution of a problem and that for another. Philosophers who reject the possibility of such inductive reasoning can allow no room for rational technology. Why else should we think it reasonable to accept that one particular medicine is safer, or more likely to be successful, than another in the treatment of a particular disease, unless there is inductively relevant evidence for thinking this? Anyone who seriously and consistently rejected the possibility of induction would be obstructing the path of worthwhile enquiry. And the same would be true for any other feature of correct methodology. So for present purposes we need to ask whether there is a case for treating realism as such a feature. That is, does realism describe an operation – not induction but some other operation – that, on a correct conception of scientific procedure, every scientific theorist should strive to pursue? Or does instrumentalism describe some such operation? Or, alternatively, do realism and instrumentalism just offer differing ontological options? Are they just opposing analyses of what scientific theories assert, rather than opposing accounts of correct methodology?

Let us consider these different versions of what the realism-instrumentalism issue is all about, and let us take the ontological version first. What

needs to be remarked, and is often overlooked, is that ontological realism cannot avoid imputing a metaphor to any new scientific theory to which it is applied. The theory is interpreted as implying the "existence" – in nature – of unobservable, particles, forces, or whatever. But in any non-theoretical assertions of existence the normal criterion for what is to be said to exist in nature is observability. Analogously Darwin (in *The Origin of Species*, 1859) adopted "selection" as a metaphorical term for nature's non-purposive process of singling out individuals with a better chance of survival, despite the fact that in its normal use the term "selection" implies conscious purpose. Of course, metaphors die and pass into the language. This is certainly what happened to Darwin's use of 'the term "selection" . The word has long since enjoyed a dictionary entry in its biological sense as well as in its pre-Darwinian sense. But it is not at all clear how "existence" or "reality" – as a term associated with the analysis of scientific theories – can achieve this status because there are no obvious criteria for its application other than the requirement that it have the very implications that are cited by instrumentalists as the content of their rival analysis: the experimentally detectable properties of the waves, particles, forces, etc. In other words ontological realism and ontological instrumentalism seem to collapse into each other. The ontological version of the realism-instrumentalism issue dissolves the conflict.

VI

The issue comes alive again, however, if we regard it as a methodological one, not an ontological one. The realist is then claiming that the supposition of a particular type of empirically undetectable entity guides the choice of one type of empirically detectable entity rather than another as having the properties that the theory should aim to predict. The instrumentalist, on the other hand, claims that he has no need of such a hypothesis: he can make his choice without such guidance. The realist asserts that the scientist's theoretical assumptions suggest the experiments by which he can test his theory: the instrumentalist asserts that the theory is no more than a set of possible experiments. And if you think that realism, thus conceived, reflects an important feature of scientific methodology, viz, the importance of developing a good basis for ongoing imaginativeness and originality on relevant issues, then that is at least one good reason to be a realist.

Someone may object at this point that there is an *ignoratio elenchi* here. The paper began by asking whether there are any *ethical*, as distinct from methodological reasons for being, or not being a scientific realist, and it has now moved away from that topic to discuss whether there are any *methodological* reasons for being or not being one.

However, if you have a general moral duty to follow the best methodological principles available and one such principle is the principle P, then normally (i.e. unless there is a conflicting principle that has priority in your situation) you must have a derivative moral duty to follow P. More specifically, if realism is part of the best scientific methodology and if your professional ethics requires you to follow the best scientific methodology, then you have a professional duty to be a realist. There is a parallel here in jurisprudence. If you have a general moral duty to conform to the law of your community, and you have a legal duty to file your tax-return each year before October 1, then you have a derivative moral duty to file your tax-return each year before October 1. Being a virtuous, law-abiding citizen means that you will conform to all applicable laws, and if you do not so conform you are doing something that is morally as well as legally wrong.

It may seem, therefore, that we have tracked down here one genuinely ethical reason for always being a realist in one's attitude to scientific theory. But there are at least two arguments for not being overmuch impressed by this achievement.

The first is that the same line of reasoning may be pursued in relation to any other correct methodological principle. So, disappointingly, there is nothing special here about scientific realism.

Secondly, and perhaps more interestingly, the attribution of a moral duty here may seem important only because it exploits an *a priori* connection. If you have a moral principle in virtue of which you ought to do act A, and if A is done if and only if B is done, then you have a moral principle in virtue of which you ought to do act B. But the ethical obligation to follow a correct methodological principle here is not inherent in the principle itself even though a priori derivable from it.

VII

Let me now sum up. I began (I) by posing the question whether there are any connections between epistemological and ethical issues about science. I then (II) pointed out that, while moral judgements in science normally involve applying some general moral principle to particular circumstances, no such application takes place in the assertion of scientific realism. Next I considered (III) what it would be like to consider science as an end in itself, as art is sometimes regarded, and (IV) how strong is the case for taking this view of scientific motivation instead of an instrumentalist one. I then (V-VI) went on to consider two different versions of what is at stake in the controversy between instrumentalism and realism, namely the ontological version and the methodological one, and pointed out that a scientist's professional ethics, by requiring him to

adopt the best available methodology, transforms each methodological principle into an ethical obligation.

Boston Studies in the Philosophy of Science

193. R.S. Cohen, M. Horne and J. Stachel (eds.): *Experimental Metaphysics.* Quantum Mechanical Studies for Abner Shimony, Volume One. 1997 ISBN 0-7923-4452-9
194. R.S. Cohen, M. Horne and J. Stachel (eds.): *Potentiality, Entanglement and Passion-at-a-Distance.* Quantum Mechanical Studies for Abner Shimony, Volume Two. 1997 ISBN 0-7923-4453-7; Set 0-7923-4454-5
195. R.S. Cohen and A.I. Tauber (eds.): *Philosophies of Nature: The Human Dimension.* 1997 ISBN 0-7923-4579-7
196. M. Otte and M. Panza (eds.): *Analysis and Synthesis in Mathematics.* History and Philosophy. 1997 ISBN 0-7923-4570-3
197. A. Denkel: *The Natural Background of Meaning.* 1999 ISBN 0-7923-5331-5
198. D. Baird, R.I.G. Hughes and A. Nordmann (eds.): *Heinrich Hertz: Classical Physicist, Modern Philosopher.* 1999 ISBN 0-7923-4653-X
199. A. Franklin: *Can That be Right?* Essays on Experiment, Evidence, and Science. 1999 ISBN 0-7923-5464-8
200. D. Raven, W. Krohn and R.S. Cohen (eds.): *The Social Origins of Modern Science.* 2000 ISBN 0-7923-6457-0
201. Reserved
202. Reserved
203. B. Babich and R.S. Cohen (eds.): *Nietzsche, Theories of Knowledge, and Critical Theory.* Nietzsche and the Sciences I. 1999 ISBN 0-7923-5742-6
204. B. Babich and R.S. Cohen (eds.): *Nietzsche, Epistemology, and Philosophy of Science.* Nietzsche and the Science II. 1999 ISBN 0-7923-5743-4
205. R. Hooykaas: *Fact, Faith and Fiction in the Development of Science.* The Gifford Lectures given in the University of St Andrews 1976. 1999 ISBN 0-7923-5774-4
206. M. Fehér, O. Kiss and L. Ropolyi (eds.): *Hermeneutics and Science.* 1999 ISBN 0-7923-5798-1
207. R.M. MacLeod (ed.): *Science and the Pacific War.* Science and Survival in the Pacific, 1939-1945. 1999 ISBN 0-7923-5851-1
208. I. Hanzel: *The Concept of Scientific Law in the Philosophy of Science and Epistemology.* A Study of Theoretical Reason. 1999 ISBN 0-7923-5852-X
209. G. Helm; R.J. Deltete (ed./transl.): *The Historical Development of Energetics.* 1999 ISBN 0-7923-5874-0
210. A. Orenstein and P. Kotatko (eds.): *Knowledge, Language and Logic.* Questions for Quine. 1999 ISBN 0-7923-5986-0
211. R.S. Cohen and H. Levine (eds.): *Maimonides and the Sciences.* 2000 ISBN 0-7923-6053-2
212. H. Gourko, D.I. Williamson and A.I. Tauber (eds.): *The Evolutionary Biology Papers of Elie Metchnikoff.* 2000 ISBN 0-7923-6067-2
213. S. D'Agostino: *A History of the Ideas of Theoretical Physics.* Essays on the Nineteenth and Twentieth Century Physics. 2000 ISBN 0-7923-6094-X
214. S. Lelas: *Science and Modernity.* Toward An Integral Theory of Science. 2000 ISBN 0-7923-6303-5
215. E. Agazzi and M. Pauri (eds.): *The Reality of the Unobservable.* Observability, Unobservability and Their Impact on the Issue of Scientific Realism. 2000 ISBN 0-7923-6311-6
216. P. Hoyningen-Huene and H. Sankey (eds.): *Incommensurability and Related Matters.* 2001 ISBN 0-7923-6989-0
217. A. Nieto-Galan: *Colouring Textiles.* A History of Natural Dyestuffs in Industrial Europe. 2001 ISBN 0-7923-7022-8

Boston Studies in the Philosophy of Science

218. J. Blackmore, R. Itagaki and S. Tanaka (eds.): *Ernst Mach's Vienna 1895–1930.* Or Phenomenalism as Philosophy of Science. 2001 ISBN 0-7923-7122-4
219. R. Vihalemm (ed.): *Estonian Studies in the History and Philosophy of Science.* 2001 ISBN 0-7923-7189-5
220. W. Lefèvre (ed.): *Between Leibniz, Newton, and Kant.* Philosophy and Science in the Eighteenth Century. 2001 ISBN 0-7923-7198-4
221. T.F. Glick, M.Á. Puig-Samper and R. Ruiz (eds.): *The Reception of Darwinism in the Iberian World.* Spain, Spanish America and Brazil. 2001 ISBN 1-4020-0082-0
222. U. Klein (ed.): *Tools and Modes of Representation in the Laboratory Sciences.* 2001 ISBN 1-4020-0100-2
223. P. Duhem: *Mixture and Chemical Combination.* And Related Essays. Edited and translated, with an introduction, by Paul Needham. 2002 ISBN 1-4020-0232-7
224. J.C. Boudri: *What was Mechanical about Mechanics.* The Concept of Force Betweem Metaphysics and Mechanics from Newton to Lagrange. 2002 ISBN 1-4020-0233-5
225. B.E. Babich (ed.): *Hermeneutic Philosophy of Science, Van Gogh's Eyes, and God.* Essays in Honor of Patrick A. Heelan, S.J. 2002 ISBN 1-4020-0234-3
226. D.D. Villemaire: *The Metaphysical Foundations of Modern Physical Science and E.A. Burtt: Historian and Philosopher.* 2002 ISBN 1-4020-0428-1
227. L.J. Cohen: *Knowledge and Language.* Selected Essays of L. Jonathan Cohen. Edited and with an introduction by James Logue. 2002 ISBN 1-4020-0474-5

Also of interest:

R.S. Cohen and M.W. Wartofsky (eds.): *A Portrait of Twenty-Five Years Boston Colloquia for the Philosophy of Science, 1960-1985.* 1985 ISBN Pb 90-277-1971-3

Previous volumes are still available.

KLUWER ACADEMIC PUBLISHERS – DORDRECHT / BOSTON / LONDON