

**Edited by**

**Jonathan M. Golding**

**Colin M. MacLeod**

**Intentional**  
**FORGETTING**

**Interdisciplinary Approaches**

**INTENTIONAL FORGETTING:  
INTERDISCIPLINARY  
APPROACHES**



# INTENTIONAL FORGETTING: INTERDISCIPLINARY APPROACHES

Edited by

**Jonathan M. Golding**

*University of Kentucky*

**Colin M. MacLeod**

*University of Toronto at Scarborough*



**LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS**  
1998 Mahwah, New Jersey London

Copyright © 1998 by Lawrence Erlbaum Associates, Inc.

All rights reserved. No part of the book may be reproduced in any form, by photostat, microform, retrieval system, or any other means, without the prior written consent of the publisher.

Lawrence Erlbaum Associates, Inc., Publishers  
10 Industrial Avenue  
Mahwah, NJ 07430

**Library of Congress Cataloging in Publication Data**

Intentional forgetting: interdisciplinary approaches / edited by  
Jonathan M. Golding, Colin M. MacLeod.

p. cm.

Includes bibliographical references and index

ISBN 0-8058-2211-9 (cloth : alk. paper). — ISBN 0-8058-2212-7 (pbk. : alk. paper)

1. Human information processing. 2. Selectivity (Psychology). 3. Memory. 4. Human information processing. I. Golding, Jonathan M. II. MacLeod, Colin M.

BF378.S45158 1998

153.1'25—dc21

97-10293

CIP

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*For our parents,  
Samuel & Eleanor Golding and Clary & Anne MacLeod,  
and for our graduate supervisors,  
Janice M. Keenan and Thomas O. Nelson,  
with thanks for everything they taught us.*



# CONTENTS

Preface	ix
1 Directed Forgetting <i>Colin M. MacLeod</i>	1
2 There's More to Intentional Forgetting Than Directed Forgetting: An Integrative Review <i>Jonathan M. Golding and Debra L. Long</i>	59
3 Varieties of Goal-Directed Forgetting <i>Elizabeth Ligon Bjork, Robert A. Bjork, and Michael C. Anderson</i>	103
4 Directed Forgetting: A Contrast of Methods and Interpretations <i>Barbara H. Basden and David R. Basden</i>	139
5 Directed Forgetting and Rehearsal on Direct and Indirect Memory Tests <i>Scott W. Allen and John R. Vokey</i>	173
6 An Illusion of Retrieval Inhibition: Directed Forgetting and Implicit Memory <i>Jerry Hauselt</i>	197
7 Disregarding Information in Text <i>Hollyn M. Johnson</i>	219
8 Directed Forgetting in Pigeons <i>Douglas S. Grant</i>	239
	vii

---

9	A Critical Analysis of Directed-Forgetting Research in Animals <i>Thomas R. Zentall, Karen L. Roper, Daren H. Kaiser, and Lou M. Sherburne</i>	265
10	Consequences of Attempts to Disregard Social Information <i>Linda M. Isbell, Heidi L. Smith, and Robert S. Wyer, Jr.</i>	289
11	Suspicion and Discounting: Ignoring Invalid Information in an Uncertain Environment <i>Yaacov Schul and Eugene Bernstein</i>	321
12	Disregarding Social Stereotypes: Implications for Memory, Judgment, and Behavior <i>Galen V. Bodenhausen, C. Neil Macrae, and Alan B. Milne</i>	349
13	Instructions to Disregard Potentially Useful Unshared Information in a Group Context <i>Jonathan M. Golding, Alan L. Ellis, Jerry Hauselt, and Sandra A. Sego</i>	369
14	Intentional Forgetting and Clinical Disorders <i>Marylene Cloitre</i>	395
15	Instructions to Disregard and the Jury: Curative and Paradoxical Effects <i>Saul M. Kassin and Christina A. Studebaker</i>	413
16	"The Jury Will Disregard . . .": A Brief Guide to Inadmissible Evidence <i>William C. Thompson and Juliana Fuqua</i>	435
17	Intentional Forgetting in Perspective: Comments, Conjectures, and Some Directed Remembering <i>Robert A. Bjork</i>	453
	Author Index	483
	Subject Index	495

## PREFACE

This volume explores the psychological phenomenon of intentional forgetting. For both of us, the first research project we ever undertook in graduate school examined directed forgetting, and it was our continuing interest in this phenomenon that led to our research collaboration and friendship. The book grew out of our desire to understand the breadth of this phenomenon, which has tendrils into surprisingly many areas of psychological inquiry. The 17 chapters examine the effect on memory of instructions to forget, think of as false, disregard, discredit, discount, not think about, or ignore information in a wide variety of contexts. These contexts range from the animal laboratory to the courtroom, with the methodologies employed in each domain varying dramatically.

Research on intentional forgetting has been conducted in various forms and under various names for at least 30 years, but until now no effort has been made to present these different perspectives in one place. That is our main goal: Using both review chapters and new empirical studies, this book brings together the many research paradigms investigating intentional forgetting, highlighting the commonalities that link these seemingly disparate areas of research. We think of the book as a sort of “case study” of one phenomenon in memory—the intention to forget, or to modify memory.

Why is research on intentional forgetting important? Such work helps to increase our understanding of how memory functions, especially with regard to its updating. In William James’ (1890) “blooming, buzzing confusion” (p. 488), we frequently are unable to process adequately all of the informa-

tion that we experience; online forgetting of some information is necessary. Moreover, we must often replace existing information with new information, as when someone we know relocates and acquires a new address and telephone number. Investigating this updating ability has been the main thrust of research on intentional forgetting, specifically those studies on the directed forgetting effect (see the opening review chapter of this volume by MacLeod). As you will read, cognitive experiments on directed forgetting have shown that we are able to deal more effectively with large amounts of information by following instructions to treat some of the information as “to be forgotten.” In this way, interference is reduced and we are able to devote all of our resources to the remaining to-be-remembered information. The mechanisms (e.g., selective rehearsal, retrieval inhibition) that lead to this reduction continue to promote new experiments, but more than a quarter century of research leaves no doubt that the directed-forgetting effect is robust.

The interest in memory updating is also partly a function of the fact that even when we can adequately process the information presented to us, there remain instances when specific information should not be processed. The most typical example of this is an instruction to disregard inadmissible evidence presented in a courtroom. The aim of an instruction to disregard is to inform jurors that this particular evidence should not be incorporated into memory or used in rendering a verdict. There are numerous other related experiences (see Golding & Long’s review chapter in this volume), ranging from advertisements that convey incorrect or deceptive information to behavioral descriptions that should be kept confidential. Despite the intent of the instructions in these studies—and unlike the findings in studies of directed forgetting—much of this research has shown that we are apt not to heed the instruction to disregard, and instead to use to-be-disregarded information to make a “contaminated” judgment.

The organization of the book reflects both of the previous points regarding the importance of research on intentional forgetting. After the directed-forgetting review (chap. 1) and the review on related intentional-forgetting phenomena (chap. 2), the book is divided into several separable but related sections. The first section includes chapters that describe recent research on directed forgetting in humans. Chapters 3 through 6 examine the mechanisms responsible for directed forgetting (i.e., selective rehearsal vs. retrieval inhibition); chapter 7 presents research on disregarding information in text. The second section (chaps. 8 & 9) describes research on directed forgetting in animals. The third section describes research on social and clinical psychological phenomena, at both the individual (chaps. 10, 11, & 12) and group levels (chap. 13), ending with a treatment (chap. 14) of recent clinical work. The fourth section focuses on psychology and law, presenting both a research overview (chap. 15) and a legal guide (chap. 16). The book

concludes with a chapter (chap. 17) by Robert Bjork, putting research on intentional forgetting into perspective.

As with any book, we owe a debt of gratitude to many people for their contributions, and we are happy to acknowledge them. First and foremost, we thank the authors of the chapters for their enthusiasm regarding the project and for hewing closer to our production schedule than we had any right to expect. We especially thank Robert Bjork for writing the final overview chapter, a wonderfully personal account from the one researcher who has contributed continuously to the study of intentional forgetting from its earliest days. We are also grateful to Rebecca Polley Sanchez for the fine work she did in constructing the indexes. Finally, we owe a considerable debt of gratitude to the team at Lawrence Erlbaum Associates—most notably our Acquisitions Editor, Judith Amsel, her Editorial Assistant, Kathy Dolan, and our Book Production Editor, Larry Hayden—for an excellent job in making this book a reality. The contributions of all of these colleagues have helped to make editing this volume an even more edifying and enjoyable activity for us.

Jonathan M. Golding  
Colin M. MacLeod

## REFERENCE

James, W. (1890). *Principles of psychology* (Vol. 1). New York: Holt.



CHAPTER

1

DIRECTED FORGETTING

Colin M. MacLeod

University of Toronto at Scarborough

When people experience failures of memory, they often express the desire for a panacea to prevent such failures from ever happening again. Of course, what they usually want is a “memory pill” to eliminate the need to exert any effort toward developing this vastly improved memory. Most people apparently believe that memory improvement should be effortless, despite recognizing that improvement of other skills such as squash or bridge requires intensive practice and dedication. Moreover, they do not seem to consider the negative consequences of never forgetting anything. We do not want to remember the loss of family and friends, the embarrassing moments, or indeed the old telephone numbers of our personal past. We need to forget, just as we need to remember.

The value of forgetting has long been understood by those who study memory. Ribot (1882, p. 61) said that “Forgetfulness, except in certain cases, is not a disease of memory, but a condition of its health and life.” Similarly, James (1890, p. 680) wrote that “if we remembered everything, we should on most occasions be as ill off as if we remembered nothing.”<sup>1</sup> Much earlier, Cicero talked in *De Oratore* of a professor offering to teach the skill of mnemonics to the very wise Themistocles. “The professor asserted that it would enable him to remember everything; and Themistocles replied that he would be doing him a greater kindness if he taught him to forget what

---

<sup>1</sup>These two quotations from Ribot (1882) and James (1890) were initially cited in the introduction to an article by Geiselman (1977, p. 323).

he wanted than if he taught him to remember” (cited in Herrmann & Chaffin, 1988). Forgetting is important to successful remembering. This chapter is about one aspect of forgetting: the forgetting that we intend to do.

## INTRODUCTION

In 1954, John Brown published an article in which he used an unusual procedure to study interference in short-term retention. Had he called his manipulation *directed forgetting*, this article probably would be highly cited today. Instead, it is essentially unknown.<sup>2</sup> Here is what Brown did in his first experiment. Subjects saw four rapidly presented arrow-number pairs on each trial. Either before or just after the stimuli were presented, instructions were given to recall only the arrows, only the numbers, or both, in a specified order. Recall of the arrows was unaffected by condition. For the numbers, however, being told in advance what to recall was much better: This was true for numbers only (.78 before, .51 after), for numbers first (.61 before, .36 after), and for numbers second (.48 before, .39 after).<sup>3</sup> What is especially noteworthy is that instructions to recall only the numbers led to better memory even when the instructions were given *after* encoding. Brown (1954, p. 147) concluded that “Some of the difference may be due to selective rehearsal of the numbers during presentation” but believed this to be an unlikely account (given the rapid presentation).

A decade later, Muther (1965) employed a similar manipulation: Subjects saw several lists of 20 letters but were told that, for some lists, only 10 of the letters would have to be recalled later. The 10 to-be-omitted letters were cued by either a preceding or a succeeding blank field. Compared to control lists of only 10 letters, where recall accuracy was .74, accuracy for cued lists averaged .61, a significant reduction. Thus, the letters that did not have to be recalled still interfered with those that did. This interference was not surprising. More interesting, however, was the comparison to 20-letter lists with no cuing. Here, correct recall was only .46. Quite clearly, being allowed to omit half of the letters from recall resulted in significantly improved performance on the remaining ones.

From these modest beginnings, a new technique arose for studying what Bjork (1972) soon characterized as the “updating of memory.” Initially known

---

<sup>2</sup>Intriguingly, this is the same John Brown who reported his research on forgetting from short-term memory (Brown, 1958) a year before the better known work of Peterson and Peterson (1959).

<sup>3</sup>In reporting data, the values in parentheses are always mean proportions correct, except in the very few cases indicated otherwise. I have occasionally collapsed over means, and have sometimes had to estimate values from figures. Hopefully, this does not do too much violence to the actual findings.

as instructions to forget (e.g., Bjork, LaBerge, & Legrand, 1968), and variously called positive, voluntary, motivated, or intentional forgetting, the name investigators eventually settled on was *directed forgetting* (see the early literature reviews by Bjork, 1972; Epstein, 1972), arguing that it was the most general (and the least theory-dependent) label. Oddly, there were no further reviews until Johnson's (1994), itself fairly selective (see also Bjork, 1978, 1989, for extended discussions). The aim of this chapter is to provide a comprehensive overview of the memory literature that has addressed directed forgetting in the past 30 years. The next chapter (Golding & Long) surveys the rest of the relevant literature outside memory research where intentional forgetting has been studied.

Here is my plan. First, I examine the studies of directed forgetting, organizing them as shown in Table 1.1. I attempt to extract the most fundamental and reliable findings to provide a clear set of phenomena to be explained. Often, I present short summaries of actual data in the form of mean proportion correct in each condition, hopefully providing more archival value. The studies are presented largely in chronological order, in this case a reasonable sequence with the advantage of making individual studies easy to locate. Once the literature is covered, I examine the existing theories against

TABLE 1.1  
Organization of the Chapter

<i>Section</i>	<i>Page</i>
Introduction	2
The Basic Paradigm	4
Oft-Forgotten Antecedents	5
The Golden Age: 1968–1974	6
Bjork, LaBerge, and Legrand (1968)	6
Bjork and Colleagues: Segregation and Selective Rehearsal	7
Epstein and Colleagues: Segregation and Selective Search	11
Elmes and Colleagues: Segregation and Short-Term Memory	14
Weiner and Colleagues: Repression	15
Other Short-Term Memory Studies	18
Other Long-Term Memory Studies	22
Summary of the Golden Age	26
The Silver Age: 1975–1985	26
Last of the Short-Term Memory Studies	27
The Growth of Long-Term Memory Studies	28
Geiselman and Colleagues: From Selective Rehearsal to Repression	34
Summary of the Silver Age	40
The Modern Age: 1986–1995	41
Directed Forgetting Confronts Implicit Remembering	41
Studies Using Traditional Explicit Testing Methods	45
The Big Picture	48
The Principal Findings and Explanations	48
The Past and the Future	52

this database. My goal is to arrive at the best explanation currently possible for the existing set of data. But first, as a “start vector,” I introduce the rudiments of the technique and define the key terms used in this chapter.

## THE BASIC PARADIGM

Consider a simple prototype experiment. My only shorthand will be to call the to-be-remembered material the *R items* and the to-be-forgotten material the *F items*, with the instruction referred to as the “cue.” In this minimalist experiment, the studied material consists of a list of unrelated words. However, the technique is applicable to a wide variety of materials and situations, as the experiments reviewed here—and in the other chapters in this book—ably demonstrate.

The Muther (1965) study illustrates one basic procedure for conducting a directed-forgetting study. There are three conditions. The directed-forgetting condition consists of equal numbers of R and F words, say 12 of each. This condition is “bracketed” by conditions representing the two logical extremes; specifically, a list of only the 12 R words, and a list of 24 words in which all items are to be remembered. In the critical list containing F cues, the subject is told that only the R items will ever be tested, though more often than not experimenters break this promise.

If the subject is instructed to forget 12 of 24 words and edits those words out of study and recall, then recall ought to be better on the remaining 12 R words than on the corresponding 12 words in a list where all 24 must be remembered. Why? There are many possible reasons, including lowered attentional demands, reduced proactive interference, more focused rehearsal, easier organization, and the like, all of which I consider along the way. Note, however, that unless the memory editing is perfectly effective, the 12 R words from the list of 24 should not be as well remembered as the 12 words from a list of only 12. That is exactly what Muther (1965) found.

How does one administer a cue to forget? There have been two basic tactics. The *item procedure*, still the most prevalent, is similar to the one Muther used. The words are presented one at a time at a fixed rate. For each F item, the subject is given an explicit cue, such as “forget” or “FFFF.” As well, researchers usually signal each R item with a complementary cue, such as “remember” or “RRRR.” Occasionally, the cue is presented before the word to which it applies, but generally the cue comes after the relevant word, ensuring that the subject registers the word.

The other procedure for delivering the cue is called the *list procedure*. This involves only one cue, ordinarily presented at the middle of the list. In my example, the subject would be presented with the first 12 words and then told either to remember those 12 words or to forget them. The remain-

ing 12 words would then be presented, always to be remembered. Variations between these two procedures are possible, such as presenting a few words before each cue (e.g., MacLeod, 1975), but these are less commonly used.

It is common to refer to a “directed-forgetting effect” when the instructional cues affect memory. However, some potential for confusion exists because there are two logical candidate measures of the basic “directed-forgetting effect.” One compares the critical 12 R words across the three conditions. I describe such effects in terms of a *cost/benefit* analysis. To the extent that the critical 12 R words do better when coupled with an equal number of F words as opposed to an equal number of R words, then the *directed-forgetting benefit* is large. To the extent that the critical 12 R words do worse when coupled with an equal number of F words as opposed to being presented alone in a 12-word list, then the *directed-forgetting cost* is large. So, the “directed-forgetting effect” can be measured by treating the difference in performance between 12 R words presented alone (the cost baseline) and 24 R words (the benefit baseline) as the entire range, and then determining where performance on the 12 R words in the critical condition (i.e., 12 R + 12 F) lies in that range. Benefit plus cost must always sum to the difference between the two end points of the range.

This cost/benefit analysis considers only the R items. Yet we generally look at performance on the F words as well (much to the subject’s consternation). Thus, in the critical condition, one can also directly examine the difference in performance between the 12 R and the 12 F words as another measure of the “directed-forgetting effect,” common in studies where the two baseline conditions are omitted. Indeed, this has come to be the preferred measure. To the extent that the F words are successfully edited out, then the difference between the R and F items should be large. I refer to this difference as the *remember–forget (R–F) difference*. Introduction of these new terms may help to reduce the confusion of labeling two different phenomena as the “directed-forgetting effect.”

This thumbnail sketch characterizes the directed-forgetting paradigm. The only other feature to note is that the dependent variable is virtually always accuracy, typically the proportion of items correct on a memory test that the subject had been told would occur. Occasionally, response latency has been recorded, but exploration of other dependent variables has been minimal. With this sketch in mind, I now review the literature.

## OFT-FORGOTTEN ANTECEDENTS

Investigators thought about lightening a subject’s memory load “on line” before the directed-forgetting procedure became prevalent (e.g., Brown, 1954). In particular, there was a longstanding interest in “repression” (for reviews, see MacKinnon & Dukes, 1962; Weiner, 1966; for experimental work,

see Glucksberg & King, 1967; Zeller, 1950). Could subjects prevent specific information in memory from becoming conscious and, if so, what were the consequences for remembering other information? Some of the relevant repression literature is examined in this chapter; some is covered in the next chapter by Golding and Long.

The idea of controlling remembering was in the air in memory research when directed forgetting was born, no doubt due to the emergence of the two-state memory models (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) and the resulting concern with transfer from short-term memory to long-term memory. This can be seen most plainly in the work on the monitoring and control of rehearsal (e.g., Atkinson & Shiffrin, 1971). Investigators were intrigued by what happened when material received differing amounts of rehearsal either freely under a person's own control (e.g., Rundus, 1971) or as instructed by the experimenter (e.g., Hellyer, 1962).

There were other relevant manipulations as well, the most directly pertinent being part-list cuing (Slamecka, 1968; for a review, see Nickerson, 1984). Here, subjects studied a list and then were given part of that list at the time of the retention test and asked to remember the rest. Remembering was poorer on the rest of the list when subjects were given the part-list cues than when they were left on their own. This cost due to the part-list cues is reminiscent of the cost due to F items in the list. But over the last 30 years, directed forgetting has created its own niche in memory research. Let us begin by examining the seminal studies in the late 1960s and early 1970s.

## **THE GOLDEN AGE: 1968–1974**

This first epoch of directed-forgetting research begins with the study that really established directed forgetting (Bjork et al., 1968). Because a few investigators were predominant in this early work, I organize my discussion in part by researcher rather than by topic, although the two overlap. As these studies heavily influenced what was to follow, I consider them in some detail.

### **Bjork, LaBerge, and Legrand (1968)**

Although they cited the Muther study, it was Bjork et al. (1968) who really brought directed forgetting to the attention of memory researchers. They reported a single experiment. Subjects were presented with 48 lists of digits, each list also containing one or two target consonant strings (e.g., BKNR). Subjects were to name the digits and consonant targets, and then to recall them after each list. In some lists, subjects were warned just before the second consonant string target that the first one could be forgotten. When both targets had to be remembered, only about .15 of second targets were

recalled, whereas approximately .39 were recalled from lists with only one target. When the first of two could be forgotten, recall accuracy for the second was about .30, a pattern like Muther's.

Bjork et al. concluded that allowing people to forget the first of two items reduced the proactive interference on the second one, improving its recall. In a way, this was another technique for producing "release from proactive interference" (cf. Wickens, 1970). Indeed, they even suggested that the case where the first item could be forgotten was more like that of a single item than of two items. They called their procedure "instructions to forget."

Most important, Bjork et al. set out four possible explanations. First, subjects might be able to *actively erase* items from short-term memory; like Muther, though, they saw this possibility as "intriguing but highly unlikely" (p. 56). Muther reported that F letters appeared more often as intrusions than did unrepresented letters, indicating that the F items could not have been deleted. Thus, the erasure view never found favor. Second, Bjork et al. thought that not having to remember the first item might lead subjects to rehearse the second item more. Because their rapid presentation was intended to minimize rehearsal opportunities, they also saw this *selective rehearsal* account as "unlikely." Interestingly, Bjork (1972) would come to prefer this explanation in his review. Both of these accounts emphasized encoding differences between R and F words.

Their third account seems to have been a sort of *output interference* idea, with the extra load of other R items impairing recall of a target item when the whole list had to be remembered. Here, the analogy to part-list cuing resurfaces. This load could at least be reduced when part of the list was to be forgotten. Such an explanation segues nicely into their final account, the one Muther (1965) preferred. Subjects might be able to segregate or partition R items from F items by "actively tagging" them, the precursor to the *set differentiation-selective search* hypothesis that Epstein (1972) favored in his review. These two approaches shift the emphasis away from encoding and toward retrieval. The stage was set for the testing of these hypotheses to begin.

### **Bjork and Colleagues: Segregation and Selective Rehearsal**

From the beginning, Robert Bjork emphasized the utility of directed forgetting as a sort of "model" (in the sense used in the animal literature) of the updating of memory. How we deal with new information that renders old information irrelevant is one of the major insights that he sees as coming from directed-forgetting research.

Bjork (1970) examined the reduction in proactive interference caused by a cue to forget prior material. Subjects studied 64 lists, each made up of one to eight syllable-word pairs presented on colored slides. Some subjects were told that, when the color of slides changed (once or not at all in a list), they

could forget all prior pairs in that list. Other subjects were not told the significance of the cue. The number of precue pairs ranged from zero to three, thereby varying proactive interference.

For subjects not told to forget, performance on postcue pairs dropped as the number of precue pairs increased, demonstrating increasing proactive interference. Furthermore, precue pairs intruded on the recall of postcue pairs. For subjects given a forget cue, the number of F pairs prior to the cue did not matter and precue F pairs rarely intruded. Thus, F pairs caused virtually no proactive interference. Recall of F pairs (measured on the last four lists only) was poor but not zero. Thus, although they caused no proactive interference, F items were not entirely forgotten.

In Bjork's third experiment, the F cue could occur either before or after the R items. He used four-pair lists with one cue in the middle and one at the end. Consider recall of the second two pairs. When told to remember both sets, subjects recalled about .57; the same was true when they were initially told to remember and then later told to forget the first set (.62). But when told to forget the first set and remember the second, recall increased (.89), a clear directed-forgetting benefit. Now consider first-set recall. When both sets had to be remembered, recall was .54; when the first set was to be remembered and the second set could be forgotten, recall improved to about .65. Intrusion results also told a consistent story: Forgetting some items helped in remembering the others.

Bjork argued that subjects used F cues in two ways: to segregate R from F items (set differentiation) and to then rehearse only R items (selective rehearsal). He saw these two processes as demanding each other in that keeping sets differentiated would require different processing, and different processing would tend to keep sets separated. This view explained why the cue first to remember and later to forget the initial set failed in Experiment 3: Subjects had grouped the sets, but had not differentially rehearsed them.

Bjork's account also explained the findings of a related study by Reitman, Malin, Bjork, and Higman (1973).<sup>4</sup> They occasionally tested F pairs and showed two things: (a) Although F pairs did not intrude or interfere with recall of R pairs, they were quite often present in memory, and (b) F pairs did interfere with each other. When the F pair immediately before the cue to forget was preceded by no other F pairs, it was recalled .54 of the time; this value fell to .49 for one prior F pair and to .24 for two prior F pairs. Bjork saw this study as evidence that selective rehearsal alone could not explain directed forgetting. The F and R items appeared to behave as quite independent sets.

---

<sup>4</sup>This research actually was carried out earlier and was frequently cited as a 1971 technical report (see e.g., Bjork, 1972; Epstein, 1972). The reason this is worthy of note is that, even in its early form, this paper bore the title "Strategy control and directed forgetting," making it the first to introduce what would become the common name for the paradigm and the phenomenon.

Woodward and Bjork (1971) switched to longer lists and used the item procedure (colored dots) to signal the cue after each word. Subjects studied six 24-word lists, attempting to recall just the R words after each list. They earned 1¢ for each R word recalled and lost 1¢ for each F word recalled. Subjects recalled about .50 of R words, intruding only about .02 of F words. After all six lists, subjects were given a final recall test and earned 1¢ for every word recalled—whether F or R. Here, they recalled .23 of R words and .05 of F words. Manipulating presentation time from 1 to 4 sec had little effect, probably because subjects waited for the cue before processing an item. When asked to circle F words on a list of all studied words, subjects correctly circled .45 of the F words and only incorrectly circled .06 of the R words. This was the first use of a measure of cue retention.

In Experiment 2, each 24-word list was composed of six 4-word semantic categories. Each category could be all R, all F, or a mixture (cf. Horton & Petruk, 1980). Categorization had an interesting effect. As the number of R words in a category increased, recall of *both* R and F words improved: For the case of 2 R and 2 F words in a category, as recall of R words rose from 0 to 1 to 2 words, recall of F words rose from .03 to .46 to .54. Finally, cue retention was better in Experiment 2, with .74 of F words correctly identified and only .08 of R words incorrectly identified.

Woodward and Bjork (1971) offered several conclusions. First, they were impressed by the subjects' ability to avoid intruding F items, given random cues within lists. Second, although F items did not intrude in recall, they were still present in memory to some extent. In particular, when F items were categorically related to R items, facilitating their retrieval, they were much more likely to be remembered (see also Halmiova & Sipos, 1976). They were also surprised that differential grouping of R and F words could be done even within a semantic category. They retained Bjork's emphasis on selective rehearsal, but viewed rehearsal as having its impact on retrieval, not encoding. F items did not interfere because they were segregated via rehearsal, hence relatively unretrievable.

In a follow-up study, Bjork and Woodward (1973) made one change: Only three of the six lists were tested immediately for R items (a distractor task filled the test interval), and an additional one was tested for both F and R items. Subjects recalled many R words (.60) but intruded few F words (.02) on the immediate test. Recall of F words only went up to .05 on the single list where subjects had to recall all words, so F words were not just suppressed. What is intriguing, though, is their conclusion that "the results implicate rehearsal and organizational processes *at input* rather than suppression or selective retrieval processes during output" (p. 22, italics added), in contrast to their earlier conclusion that "the differential rehearsal devoted to R words operates primarily *on retrieval* rather than on storage" (Woodward & Bjork, 1971, p. 109, italics added). This "about-face" aside, they still

favored the explanation that rehearsal led to set differentiation of the R and F items.

Woodward, Bjork, and Jongeward (1973) went on to argue that there was a type of rehearsal that simply maintained items in memory until the cue appeared. Subjects studied four 36-word lists, half R and half F items, using the item procedure. Each word was presented for 1 sec followed by a variable blank period prior to the 1-sec cue. Three R words and three F words were followed by each of six blank periods—0, 1, 2, 4, 8, or 12 sec— included to encourage differential rehearsal of the items.

The immediate and final recall data closely replicated Woodward and Bjork (1971; Bjork & Woodward, 1973). Intriguingly, the duration of the rehearsal interval prior to the cue had very little effect on immediate or final recall. Only when they measured final recognition in their third experiment did they find a beneficial effect of precue rehearsal interval duration. They concluded from these recognition data that any rehearsal, whether maintenance or elaborative, led to improved long-term retention of an item. However, only elaborative rehearsal benefited recall because it encouraged interitem association. Elaborative rehearsal was much more likely to occur for R items than for F items.

Woodward, Park, and Seebohm (1974) focused on the impact of an F cue at encoding versus retrieval. Subjects saw six 24-word lists, with one color cue designating the 8 F items in a list and two other color cues indicating the two sets of 8 R items. After each list, subjects were told what to recall. In immediate recall, two main findings emerged: (a) F words intruded very little, even when subjects were told to recall them, and (b) R words that subjects were told not to recall intruded considerably more than F words. In final recall, where subjects were to ignore previous instructions, there were again two main findings: (a) F words were still very difficult to recover, and (b) having had to recall F words immediately did not help them, but did somewhat reduce R-word recall. In final Yes–No recognition, there was a strong R–F difference, but memory for F items was evident.

Overall, Woodward et al. (1974) concluded that explicit cues to forget an item were quite different from implicit cues to not recall that item, both in recall and in recognition. They believed that it was the immediate contingency of the explicit F cue that made the difference. Subjects ceased rehearsal of explicit F items, whereas they only tried to edit recall of implicit F items. Thus “true” forget items were much less available in memory than were items simply not to be recalled.

In his review of this era, Bjork (1972) emphasized that we often need to change the priority of information in memory, and that a mechanism like directed forgetting may be one of the tactics we use to do so. His explanation of directed forgetting in this period is best put in his own words: “subjects take advantage of an F-cue in two ways: (a) They devote all further rehearsal,

mnemonic, and integrative activities exclusively to R-items, and (b) they differentially group, organize, or code R-items in a way that functionally segregates them from F-items in memory” (p. 229).

### **Epstein and Colleagues: Segregation and Selective Search**

William Epstein was the other dominant influence early on. He began by telling subjects the order in which to output items after list presentation (Epstein, 1969a), and was not surprised that recalling the beginning of a list before the end lowered recall of the end relative to recalling the end first; that was expected as a consequence of proactive and output interference. But, when told to output *only* the end part, subjects’ performance on the end part was better than when they output the end part first, followed by the beginning. The obvious difference was that the beginning could be disregarded in the “only” case, analogous to a forget cue. He decided to explore this more closely.

Epstein (1969b) constructed eight 16-item lists, half of each list consisting of two-digit numbers and half consisting of words. Subjects were tested four times, after lists 2, 4, 6, and 8. On each test, subjects were cued to recall only words, only numbers, or one before the other. The results were similar for words and numbers, so I summarize just the word data. For the second half of the list, recalling only words (.70) was better than recalling words first (.62) or second (.49). For the first half of the list, output instruction made no difference (about .50 in all three cases). Experiments 2 and 3 confirmed this pattern for longer lists and equal recall time for different cues.

Experiments 4 and 5 switched to the item procedure, randomizing words and numbers in the list. This change eliminated the advantage for the “only” condition over the “first” condition. Epstein concluded that, to be effective, directed forgetting required discrete subsets of items. He offered two explanations for the advantage of the forget (only) condition. Put in the terminology of directed forgetting,<sup>5</sup> these were: (a) When the F cue is given, the F set is deleted from short-term memory, minimizing interference, and (b) because the F half of the list need not be rehearsed, it causes minimal inhibition on retrieval of the R half. The first idea corresponds to the erasure hypothesis and Epstein, too, rejected it. But the second hypothesis, which Epstein opted for, is apparently the first mention of what would eventually become the *retrieval inhibition* explanation.

Shebilske, Wilder, and Epstein (1971) had subjects study 24 lists, each made up of four syllable–word pairs presented for 2 sec each. For half of

---

<sup>5</sup>Briefly, Epstein and his colleagues referred to directed forgetting as the “only effect” (e.g., Shebilske, Wilder, & Epstein, 1971). Unless you were a directed-forgetting researcher, though, this label might seem a little “exclusionary,” so they began using the term *directed forgetting* in 1972.

the subjects, the four words in a list were unrelated; for the other half, the first two words were from a different category than the second two. The cue was presented after presenting all four pairs and just before testing one pair. The procedure was equivalent to instructing subjects to remember all of the pairs or to forget the first two or the second two pairs. The advantage for the lists with both F and R items over those with only R items was greater for categorized (.23) than for uncategorized (.10) lists.

Intrusions were much less frequent for F items (.20 of intrusions) than for R items (.43). Also, intrusions of F items were less common in categorized lists (.12) than in uncategorized lists (.28), but there was no corresponding difference in R-item intrusion. Epstein et al. saw their findings as supporting a selective search account. In particular, distraction-filled versus unfilled intervals had little impact on the size of the directed-forgetting benefit. Filled intervals should prevent rehearsal, so this result seemed inconsistent with a selective rehearsal account. It is also noteworthy that their study showed that Bjork's findings generalized to the case of a cue given after all of the items in a set had been studied, a finding which also led Shebilske et al. to prefer the selective search account.

Epstein, Massaro, and Wilder (1972) attempted to pit the selective search and selective rehearsal accounts against each other. Their lists contained 6 pairs, with three response words from each of two categories. After each list, the cue was given and then a 3-sec filled or unfilled interval occurred. The selective rehearsal account predicts a larger directed-forgetting benefit for the unfilled intervals. Two tests were used: Half the time, subjects had to recall the response word of a critical pair; half the time, they had to pick it from the three R possibilities. Epstein et al. argued that the selection test made the search set for lists with F pairs equivalent to the search set for lists without F pairs. Therefore, directed forgetting should be evident only in recall and not in selection.

In Experiment 1, there was a directed-forgetting benefit for R words of about .13 in recall but only about .03 in selection, and this was true for both the filled and the unfilled intervals. (Note that R intrusions at .41 again vastly outnumbered F intrusions at .11.) However, when the selection test was modified in Experiment 2 to include all six possible responses—both the R and the F responses—the magnitude of the selection effect (.13) was virtually identical to that in the first experiment. Epstein et al. maintained that F cues provided tags to differentiate items at the time of retrieval, permitting subjects to search only among the R items.

Epstein and Wilder (1972) added a new wrinkle. On 8 of the 24 lists, they tested an F pair instead of an R pair, informing subjects of this on half of those trials. As always, they found higher R recall when a list contained F items (.70) than when it did not (.52). In contrast to Epstein et al., however, this advantage tended to be larger for unfilled intervals, implicating a role

for selective rehearsal. Subjects correctly recalled the F items .44 of the time when informed but only .15 of the time when uninformed. F items intruded into the recall of R items only when subjects were trying to recover another F item. Otherwise, most intrusions were other R items.

The results were seen as strong support for selective search. Two new results buttressed this claim: (a) Subjects rarely recalled an F item when they did not know that an F item was being tested, and (b) when subjects did know that they were to recall an F item, their recall of F items improved dramatically, though not to the level of R items. However, Epstein and Wilder (1972) allowed that selective rehearsal also seemed to be operating, assigning it the role of maintaining the R set's readiness to be searched, thus enhancing the effect of selective search. This is a rather different interpretation than Bjork's that both processes are necessary and working in tandem.

Instead of testing only one pair from a short list containing F items, Shebilske and Epstein (1973) sometimes unexpectedly tested all of the pairs from the list in succession. This procedure was designed to abolish the opportunity for using selective search. They reasoned that this sort of "catch trial" would not eliminate the cue effect if the relevant operations had been done prior to retrieval (i.e., if the mechanism had to do with encoding). If, on the other hand, the cue effect was largely a retrieval phenomenon due to selective search, then it should disappear because the subject could not restrict search when all items were to be tested.

As always, memory was better for R items from a list containing F items (.69) than from one containing only R items (.53). However, for trials containing F items, the directed-forgetting benefit was smaller on "test all" trials (.06) than on trials where only R items were tested (.16). As usual, there also were fewer intrusions from F items but when subjects had to recall all of the items from a list, F items made up about half of the intrusions. Because the "test all" trials virtually eliminated the cue effect, Shebilske and Epstein (1973) concluded that selective search was the key mechanism. When conditions prevented this mechanism from operating, the effect essentially vanished. In their experiments, manipulations of rehearsal opportunities did not produce such dramatic effects.

In his review, Epstein (1972, p. 187) strongly advocated the selective search hypothesis, stating that "the data are sufficient grounds for concluding that selective search can account for many of the findings of directed forgetting experiments." He noted, though, that "unequivocal positive findings of erasure have not emerged" (p. 188), and allowed that this view still might have "genuine merit." He rejected a perceptual-encoding explanation on the basis that the cues came after perception and encoding, though this does imply quite a narrow definition of encoding. Most provocative, though, was his position on rehearsal: "We consider as uninteresting those effects of forget cues that are fully accountable in terms of differential rehearsal"

(p. 151). To Epstein, the mnemonic benefits of rehearsal were not “news-worthy,” a view shared by Johnson (1994) in her recent review.

### **Elmes and Colleagues: Segregation and Short-Term Memory**

David Elmes was one of the first to study directed forgetting in short-term memory. In Elmes (1969a, Experiment 2), subjects studied 12 lists where tests of some word pairs were interspersed with presentation of the other pairs. After either 8 or 10 pairs (from lists of 12 or 14 pairs, respectively), half of the subjects knew that they could “disregard” all prior pairs, and half did not. Although not explicitly told to forget, subjects could trust the cue because precue pairs were never tested. Contrary to Bjork et al. (1968), Elmes reported no reliable recall advantage on a subsequent critical pair for the group that could forget earlier pairs (.42) compared to the group that could not forget them (.37). Elmes (1969a) argued that the “isolation or emphasis” of the single critical item in the Bjork et al. lists, not the reduction in proactive interference, was the crucial element. Elmes’ view is in essence a version of the tagging or selective search account.

Elmes (1969b) manipulated the amount of information to be forgotten. Subjects studied 24 lists of seven or eight pairs. This time, the cue was explicit and more obvious than in Elmes (1969a), and amounted to a cue to forget either three, four, or five untested prior pairs. As it turned out, the number of pairs prior to the cue did not matter, but recall of a subsequent critical pair was better following an F cue (.59) than following an R cue (.40). Elmes (1969b) held that the more explicit cue caused the stronger effect than in his prior study (Elmes, 1969a), now attributing this to selective rehearsal of the critical pair. The idea was that all items resided in short-term memory only as long as they were rehearsed, and that the F cue basically was a cue to discontinue rehearsal of the designated items, which then were not transferred to long-term memory. Tagging or isolation could not be the important feature because the same events transpired for both groups prior to the critical item.

In Elmes, Adams, and Roediger (1970), after half of the list was presented, half of the subjects were cued to forget prior pairs. Subsequent recall of the critical pair that followed the cue did not differ for the two groups. In Experiments 3 and 4, a cue was presented to both groups, with only one group told that this meant to forget all prior pairs. Now the directed-forgetting benefit reappeared in recall. In Experiment 4, the advantage was .24, with the cued group recalling .51 of critical pairs and the uncued group recalling only .27. Furthermore, intrusions from prior F items made up only .08 of all responses for cued subjects, but .25 of the responses for uncued subjects.

Elmes et al. (1970) added a four-alternative recognition test for the pairs. Of the correctly recognized items, .24 were precue items for the cued group and .26 were precue items for the uncued group, so there appeared to be

no directed-forgetting benefit in recognition despite the clear advantage in recall. Unfortunately, Elmes et al. did not report postcue performance. Nevertheless, along with the study by Gross, Barresi, and Smith (1970), described later, this appears to have been the first published use of a recognition test to measure directed forgetting.

Elmes et al. considered four possible explanations for directed forgetting. First, they argued against active erasure and selective rehearsal because recognition did not differ for R and F items, showing that the F items were still in memory. Although they considered repression of F items, they were not impressed with the support for such a view. Instead, they favored *set differentiation*, closely allied to the selective search idea. They maintained that items in the F set were suppressed whereas those in the R set were selected. The mention of suppression would appear to be, following Epstein (1969b), an early version of what would become the *retrieval inhibition* account.

The final study in this set was a change of pace. Elmes and Wilkinson (1971, Experiment 1) showed subjects lists containing a four-word category, an eight-word category, and two eight-word categories in which four words were designated as F items. In Experiment 2, an uncategorized set of eight words was added, half cued to be forgotten. Collapsing over the experiments, recall of the four-item R-only categories (.43) exceeded that for the four R items from the F-cued category (.36), as expected. But performance for the eight-item R-only categories (.48) was best of all. Oddly, then, there was a directed-forgetting cost without a corresponding benefit. This may be because scoring appears to include the first four items in the eight-item categories; a better comparison would have been to only the last four of the eight. The uncategorized items from Experiment 2 also differed (R: .28, F: .05). Finally, despite being asked not to remember the precue items in F categories, subjects still recalled .20 of the categorized F items. The size of the remember-forget difference was similar for the categorized (.16) and uncategorized (.23) items despite overall better recall for the categorized items. Once again, a subsequent recognition test on just the last list (with twice as many distractors as targets) showed no effect of cue. This recognition-recall difference, together with a preliminary analysis of clustering in recall, was taken as confirming set differentiation of R and F items, but now there was evidence that such differentiation could occur even within a category.

The Elmes studies teach us several things. First, the F cue must be distinctive, and explicit use of the word "forget" seems to make a difference. Second, it is apparently valuable to compare recognition and recall. Third, both the directed-forgetting cost/benefit and the remember-forget difference are robust (but not guaranteed) effects, at least in recall. Fourth, these effects obtain for relatively short lists as well as for longer lists; the phenomenon is not limited to short-term memory.

### Weiner and Colleagues: Repression

Because of his interest in repression (cf. Weiner, 1966), Bernard Weiner (1968) took a different tack in studying directed forgetting in short-term memory. In Experiment 1, he showed subjects letter trigrams and tested each 3, 9, or 17 sec after its presentation; number shadowing filled each retention interval. In Experiment 2, half of the subjects were asked to forget the trigrams between presentation and test on each trial whereas half were told to remember them. Forgetting rate was much more precipitous for the F group (.93 at 3 sec, .64 at 9 sec, and .40 at 17 sec) than for the R group (.97, .84, and .71, respectively). When, in Experiment 4, Weiner switched to a within-subject design and cued subjects immediately before each presentation whether to remember or forget that item, he replicated this pattern: for the F group (.87, .48, and .36) and for the R group (.93, .78, and .65). Giving the cue after presentation of the trigram also generated the same pattern in Experiment 5.

Further experiments in this series examined the situation where a shock was used to motivate forgetting. This procedure originated with Glucksberg and King (1967), who had subjects learn a 10-pair nonsense trigram-word list. Each word had a remote associate, as in the example pair *cef-stem*, where the remote associate *smell* can be reached through the sequence *stem-flower-smell*. Once subjects knew the initial pairs, they then studied a new list of pairs where the responses were the remote associates (e.g., *cef-smell*). Subjects were shocked each time 1 of 3 of the 10 pairs appeared until they could correctly anticipate those shocked pairs. Then, they had a single relearning trial on the original list. Recall of word responses from the original list was poorer after relearning if the response word's associate had been shocked (.71) than if it had not been shocked (.94). Glucksberg and King saw this as an experimental analog of repression caused by shock.

Weiner and Higgins (1969) maintained that the effects observed by Glucksberg and King were due to differential original learning, not to repression. In their Experiment 1, Weiner and Higgins failed to replicate Glucksberg and King: Forgetting was almost equivalent for both the shocked (.80) and unshocked (.82) pairs. They claimed that the pairs associated with shock in the Glucksberg and King study coincidentally were the hardest pairs to learn (although their Experiment 3 did replicate Glucksberg and King's original difference). In a reply, Glucksberg and Ornstein (1969) argued that shock and original learning level interacted, but that shock certainly affected recall.

Returning to Weiner's (1968) study, he incorporated shock for his final three experiments. In Experiment 6, he observed more forgetting for items associated with shock cues during both encoding and retrieval (.80, .57, .46 over the 3-, 9-, and 17-sec retention intervals) than for items that had an initial shock cue dispelled by a subsequent shock cue (.87, .55, .53) or for items that had no shock cue (.87, .58, .52). Weiner (1968, p. 223) maintained

that his results supported "a process of active forgetting which is attributed to differences in memory storage as opposed to differences in trace formation." He explicitly rejected differential rehearsal, and also pointed out a concern that has not always received sufficient consideration in the directed-forgetting literature: "Perhaps the most serious experimental problem . . . is differential withholding (suppression)" (p. 223). Were subjects simply being cooperative, and how often is this a potential alternative to the more exciting idea of forgetting on cue? Demand characteristics (Orne, 1962) loom large as a threat to this paradigm.

Weiner and Reed (1969) attempted to test whether directed forgetting affected encoding or retrieval in short-term memory. They followed Weiner's (1968) procedure but used three cues, delivering one just prior to each stimulus: remember and rehearse, remember but do not rehearse, and forget. Because the results were similar, I present just their within-subject version of the experiment. Over the three retention intervals (3, 9, and 17 sec), recall declined differentially: for remember-and-rehearse (.97, .88, .82), for remember without rehearsing (.95, .77, .60), and for forget (.94, .64, .51). Forgetting rate was enhanced by nonrehearsal, but even more so by an F cue. Their second experiment tackled rehearsal directly. Subjects studied each item with the goal "to remember the stimuli without rehearsing, unless signalled to forget" (p. 228). All items were tested after 13 sec, but the F versus R cue could appear at 3, 6, 9, or 13 sec. If rehearsal is crucial, delaying the cue should reduce its effectiveness. Although F items were recalled about 8–10% less than R items, there was no effect of cue timing, which Weiner and Reed saw as further evidence against differential initial learning.

Their third experiment tackled retrieval directly. Here, each of 12 trigrams was shown twice, once in set 1 (trials 1–12) and once in set 2 (trials 13–24). All possible cue combinations were used. A large overall remember–forget difference (about .25) occurred in each set, as usual. More interesting, though, was the effect of the cue staying the same or changing on individual items over the two sets. On second presentation, performance was just as high for F–R items as for R–R items, and performance was just as low for R–F items as for F–F items. Weiner and Reed argued that memory for the first presentation of an F item was available but not used unless "released" when that item became an R item on the second presentation. Retrieval of a former F item was completely disinhibited when that item became an R item, neatly presaging Bjork's (1989) argument.

Weiner and Reed's final question concerned the extent to which directed forgetting was due to suppression (conscious) or repression (unconscious). They repeated Experiment 3, interviewing subjects extensively afterward. Two groups were isolated—suppressors (one third) who consciously used differential retrieval for R and F items, and repressors (two thirds) who did not. Weiner and Reed took the very similar patterns of data in these two groups as evidence that their findings did not depend on conscious strate-

gies; I am not convinced. It seems as reasonable to argue that the demand characteristics that led subjects to withhold F items also led many of them to deny having "cheated," so that the apparent consistency in the two strategies really could reflect only one strategy.

Reed (1970) focused on retroactive interference (see also Decker, 1976). Many prior short-term memory studies gave the F cue after the first but before the second of two items, which he reasoned might have influenced encoding of the second item. So he moved the cue after the second item, following it by a 10-sec distraction interval. In Experiment 1, Reed contrasted recall of a single trigram studied alone (.91) to that when followed by an additional R trigram (.69) to that when followed by an additional F trigram (.94). The postencoding forget cue appeared to eliminate retroactive interference. In Experiment 2, Reed contrasted an F cue before versus after the filled interval and found that the "late" forget cue reduced recall from .54 to .44, the latter value being the same as in a control condition where only the first trigram was to be recalled. This suggests that a major function of a prerecall F cue is to eliminate retroactive interference.

But why does the F cue before the retention interval have an even larger effect? To discount a rehearsal explanation, Reed examined repeated presentation of the same items under changing instructions (cf. Weiner & Reed, 1969). Experiment 3 captures his idea. He ran a single trial in each of the conditions of Experiments 1 and 2 (except for the late forget condition), replicating the previous patterns. Then, all of the items from these trials were re-presented as R items. Now, recall of the trigrams did not differ as a function of original condition. Apparently, the conditions had not led to differential learning, so Reed concluded that the F cue affected retrieval from short-term memory, not storage. Interestingly, he argued that rehearsal played the role of maintaining items, not learning them, a view that would soon resurface (Woodward et al., 1973).

With the exception of Roediger and Crowder (1972; see later), Weiner was alone in his early emphasis on repression, which he also called retrieval inhibition; even his coauthor Reed did not use the term in subsequent writings. Yet repression would return in the work of Geiselman, Bjork, and Basden. Setting aside the problem of demand characteristics, the results of Weiner and Reed did seem to fit with the idea that F items and R items were equally well learned but that F items were difficult to recover. With selective search and differential rehearsal, retrieval inhibition would eventually become one of the three major explanatory mechanisms for directed forgetting.

### **Other Short-Term Memory Studies**

Turvey and Wittlinger (1969) were among the first to examine the placement of the F cue. In Experiment 1, subjects saw a trigram for 2 sec, counted backward for 1 sec, and then tried to recall. For one group, F cues appeared

at encoding, simultaneously with the trigram. For another group, all items were R items. Recall accuracy for R items in the two groups differed by about .20, favoring the group with F cues. In Experiment 2, locating the F cue at retrieval instead of encoding almost eliminated the difference between groups. Turvey and Wittlinger opted for a selective rehearsal explanation: Critical R items were rehearsed more only in Experiment 1, where rehearsal of other items could be avoided. They did admit, though, that not having to recall the items in Experiment 2 might have been crucial, recognizing the possible impact of differences in output interference.

Block's (1971) study is reminiscent of Reed's (1970). Following Bjork et al. (1968), he cued subjects to forget either the first or the second of two trigrams on some trials. When the second item was tested, subjects told to forget the first (.81) did much better than those not so instructed (.69). In fact, performance in the F-cued condition was no different from a single-item control (.84). However, when the first item was tested, subjects told to forget the second item (.46) were no better than those who had to remember it (.47), both substantially poorer than the single-item control (.76). Directed forgetting alleviated proactive interference but not retroactive interference. Because he had included a shadowing task to minimize rehearsal, Block favored a selective search or segregation account.

In Experiment 2, Block had subjects either remember or forget the first 6 words of a 12-word list; they always had to remember the last 6 words. Recall of the last 6 words was almost equivalent for subjects told to forget the first 6 (.58) as for those who did not see the first 6 (.60), but considerably worse for those told to remember all 12 words (.41). Block was among the first not only to test F items, but also to add a recognition test. The virtually identical recognition for the last 6, whether after an F cue ( $d' = .60$ ) or an R cue ( $d' = .63$ ), led him to argue against erasure and selective rehearsal because all memory tests ought to have been affected. This recognition–recall contrast came to be a standard argument in favor of the segregation explanation.

Roediger and Crowder (1972) investigated whether selective rehearsal could explain the short-term memory effects that Weiner (1968) took as evidence for repression. They thought Weiner's constant rate for his interpolated task might have prevented rehearsal only in some subjects, so they adopted the familiar Peterson and Peterson (1959) procedure. On each trial, subjects read a trigram and then counted backward from a 3-digit number for 3, 9, or 17 sec before attempting recall. A cue at encoding told subjects that, during number counting, they should either rehearse the item without disrupting number counting, remember the item without rehearsing, or forget the item. A final Yes–No recognition test followed all 54 trials. Separate counting-only blocks provided a baseline for counting performance on the critical trials.

Both main effects—cue and retention interval—were significant, as was the interaction. Thus, trigram recall was better with rehearsal (.91, .79, .78) than

without rehearsal (.80, .72, .49), and worse for F items (.70, .52, .38). The counting data concurred, with counting fastest in the F condition, intermediate in the no rehearsal condition, and slowest in the rehearsal condition. Turning to recognition, the two R conditions hardly differed (.44 for rehearsal, .41 for no rehearsal), but both were more accurate than the F condition (.33). Roediger and Crowder maintained that there was no evidence in support of a retrieval inhibition account. Rehearsal was the critical mechanism.

Spector, Laughery, and Finkelman (1973) also worried whether prior attempts to prevent rehearsal (e.g., Shebilske et al., 1971) had been successful. They used trigram-word pairs in 32 lists of either 4 or 8 pairs with or without a filled or unfilled rehearsal opportunity following the list. Subjects studied each pair for 2 sec, then, halfway through the list, a 3-sec blank occurred, then the second half of the list was presented, followed by a 3-sec rehearsal/no rehearsal period. In Experiment 1, the cue always preceded the rehearsal interval; in Experiment 2, the cue either preceded or followed the rehearsal interval, which was always unfilled. The cue could indicate a test from only the first half of the list, only the second half, or either, like the procedure of Shebilske et al. (1971).

In Experiment 1, rehearsal definitely helped cued recall, as did being allowed to recall only half of a list. For four-item lists, having to recall all words resulted in accuracies of .50 for the rehearsed and .32 for the unrehearsed pairs. Only having to recall half of the list boosted recall to .68 for rehearsed and .39 for unrehearsed pairs. A muted version of this same result was evident for eight-item lists. The directed-forgetting benefit was considerably greater for rehearsed than for unrehearsed items, and about twice as large for shorter lists as for longer lists. This is in keeping with the critical role of rehearsal because there was about twice as much opportunity to rehearse in short lists. Experiment 2 supported this claim: When the cue preceded the rehearsal interval, recall was better if only half of the list was tested (.60) than if the whole list was tested (.50); when the cue followed the rehearsal interval, this directed-forgetting benefit vanished (both conditions at .50). Making differential rehearsal impossible by delaying the cue until after the rehearsal interval ended caused all evidence of a cue effect to disappear. Once again, selective rehearsal seemed to be the key.

Timmins (1973) closely followed Bjork's (1970) procedures. Subjects studied trigram-word pairs for 4 sec each in lists of eight R pairs, four R pairs, or eight pairs with a midlist cue to forget the first four. In a novel twist, an item from the F set was sometimes repeated in the R set. For just the comparable last four R items across conditions, the standard pattern obtained: Recall was better for the four-item list (.57) than for the eight-item list (.48), but the list containing an F cue showed no cost (.59). Most important, repeated items (.67) were better recalled than all others. If set differentiation was perfect, the probability of recalling this repeated F item should

be no greater than that of the unrepeated R items because the prior instance of the repeated item was in the F set, and hence segregated. However, if rehearsal was the key, then the repeated item should be better recalled than its companion R items.

Waugh (1972; see later) previously reported a similar finding using free, rather than cued recall. Timmins argued that subjects were concentrating their rehearsal on the repeated items. But perhaps they had noticed the repeated items and directed rehearsal to them. Timmins (1975) tried to address this possibility, but a probable ceiling effect prevents straightforward interpretation.

In a close replication of Bjork et al.'s (1968) procedure, Yinger and Johnson (1973) extended the retention interval, measuring the effect of the forget cue at 6, 15, and 40 sec. They found a reliable directed-forgetting benefit at 6 sec and at 15 sec (both about .25 better when allowed to forget the first trigram vs. having to remember both), but no such difference after 40 sec. They attributed this to the short-term/long-term distinction, arguing that directed forgetting was limited to short-term retention intervals. Unfortunately, they did not explain why recall in the F condition dramatically *improved* (by about .30) from the 15-sec to the 40-sec interval, which calls their study into question.

Homa and Spieker (1974) predicted that, if selective search was operative, recall RTs should increase as the number of R items increased, and decrease in lists where an F cue allowed some items to be omitted from search. Subjects studied lists of either four, six, or eight word pairs. The use of pairs permitted timing from the onset of the stimulus term at the time of test. After half of each list was presented, a filler slide appeared, then the rest of the list was presented, followed by another filler slide, the cue, and the test slide. Items were presented for 2 sec with a 1-sec gap. Turning to the data, as predicted from the selective search view, RTs increased with list length for all lists and cue conditions, and subjects recalled R responses faster for lists containing an F cue (1.53 sec) than for all-R lists (1.73 sec). These results also fit with the accuracy data of Shebilske et al. (1971). Homa and Spieker concluded in favor of a selective search or retrieval account, but noted that this might only apply when a single F cue followed the list.

The last three studies in this section moved farther afield. Johnson (1971) used a new dependent measure, pupillary response. He gave five-word lists to subjects at a 2-sec rate. One group received midlist F cues for all prior words in that list; the other group did not. Johnson observed a highly contingent dilation-constriction pattern immediately after the F cue, with no corresponding change in the control group. Oddly, however, he did not report the recall data, so it is difficult to evaluate his finding. Furthermore, he pointed out the perennial problem with measures such as pupil response, specifically that "in many cases such pupil changes reflect the influence of information processing confounded with motivational effects" (p. 318). Al-

though the pupil pattern suggested a change in information load given an F cue, it is hard to know whether this is motivational, cognitive, or both.

Martin and Kelly (1974) closely followed up Johnson (1971). In place of pupil dilation, they used a secondary simple RT task: While doing the memory task, subjects had to press a button as quickly as possible to turn off a light. In Experiment 1, they found only a small accuracy advantage for R words in lists with an F cue (.97) versus no F cue (.94), but a ceiling effect could be occurring here. The RT data are more interesting. For all-R lists, RTs increased as encoding progressed through the five studied items, and then decreased as retrieval progressed through the five items. Subjects seemed to load up and then unload. When an F cue occurred, secondary RTs dropped sharply, much like the pupil dilation effect. Furthermore, the fewer items that had to be remembered, the faster the RT. Martin and Kelly hypothesized a reduced search set in lists containing an F cue, although the data could also reflect reduced need for resource-demanding rehearsal of F items.

Burwitz (1974) extended the domain of directed forgetting to short-term memory for simple motor responses. Subjects had to learn lever displacements without visual feedback, and then reproduce them as accurately as possible following either a short (5 sec) or long (90 sec) retention interval. In the control condition, the test followed one movement. In the R condition, four other movements preceded the critical one; at test, subjects had to reproduce the critical movement first, then the others. In the F case, the four prior movements were cued to be forgotten right before or right after the critical movement. Error in reproduction of the critical final movement was the dependent measure. At the 5-sec retention interval, the R case and the F-after case were less accurate than the F-before and the control cases. At the 90-sec retention interval, the effects magnified and spread, with accuracy of reproduction still equivalent in the control and F-before case, but both F-after and R were even less accurate. Burwitz argued that there was clearly a strong role for cognitive control of interference in motor learning, as illustrated by the effect of an F cue.

Overall, the evidence for a directed-forgetting effect in short-term memory is considerable. If subjects are permitted or encouraged to cease thinking about an item, that item causes sharply reduced interference on other items that the subject must remember. Furthermore, the F item is rather poorly recalled, although its recognition may suffer little decrement. With the decline of short-term memory studies, directed-forgetting research moved toward studies of more extended learning and remembering. This trend was already becoming evident even in this first epoch.

### **Other Long-Term Memory Studies**

This section features studies that are not all closely interrelated, but that involve directed forgetting of items in longer lists, outside the usual realm

of short-term memory. Investigators began to explore the boundary conditions and susceptibilities of the novel directed-forgetting procedure.

Bruce and Papay (1970) explored the source of primacy in free recall: Was reduced proactive interference or more rehearsal the key? Subjects studied five lists of words at 2.5 sec per word. In two lists, all words were to be remembered; one list contained 20 words and one contained 35 words. In two other lists, an F cue appeared after 15 words, then either 20 or 35 more R words were presented. For these four lists, subjects were to recall only the R words. Finally, for half of the subjects, the fifth list had 15 words before an F cue and 20 words after it, and subjects were asked to recall both R and F words. Control subjects had a fifth list containing 35 words without a cue.

Over the first four lists, primacy was equivalent for lists with versus without prior F words. Under the proactive interference account, lists with prior F items should have shown reduced primacy. Now consider the final list. For subjects given an F cue, words were recalled more poorly in positions 1–5, just as well in positions 6–11, and more poorly in positions 12–15 compared to subjects without an F cue. Apparently, rehearsal of the last few words prior to the cue—the words presumably in short-term memory—was interrupted. Perhaps, too, continued rehearsal of the earliest items was curtailed. Interestingly, these patterns were not evident on a four-alternative recognition test. Both tests did show, however, much better memory after the cue for those subjects given an F cue. Bruce and Papay replicated this pattern with an auditorily presented list (Experiment 1) and also demonstrated that simply including an odd item (the von Restorff effect; see Wallace, 1965) in the position of the F cue did not lead to the same pattern (Experiment 3). They concluded that primacy was the result of differential rehearsal of early list items.

Archer and Margolin (1970) examined whether the arousal properties of white noise would affect the action of R versus F cues. Subjects studied lists of 16 two-digit numbers, half followed randomly by each cue. Crossed with cue was the presence versus absence of a burst of 100 dB white noise before or after the list. Numbers and cues were presented for 3 sec each, with white noise for 1 sec. The test involved selecting from a complete list all those numbers that had been shown with an R versus an F cue. There was a strong effect of R (.74) versus F (.17) cues on digit identification, probably enhanced by the fact that F items were not identified until after R items. Noise did improve performance for R items (.81 with, .69 without), but had no effect on F items. They argued that the F cue terminated both further processing of the item and any advantage due to arousal.

Bugelski (1970) examined what he called “anti-intentional learning” (a term that did not catch on). Subjects read each of 32 words in a list aloud four times. Half of the words had a plus sign above them. One group was told to learn only those words with the plus sign, and to inhibit, not learn,

or “tune out” the words without the plus sign; the plus sign was not explained to the other group. The control group recalled .42 of the signed words and .45 of the unsigned words. The experimental group, who had the “forget” cue, recalled .56 of the signed words but only .26 of the unsigned words. Thus, there was a benefit of .14 for R items and a cost of .19 for F words.

Bugelski also explored imagery. Of primary interest, a group instructed to image all words but to then inhibit those without signs recalled almost the same proportion of signed (.49) and unsigned (.46) words (although it is odd that imaging conferred no advantage over the control group). A group instructed to image only the signed words and to inhibit the others showed an advantage for signed words (.61) over unsigned words (.22)—a benefit of .12 and a cost of .24. Overall, imagery instructions hardly altered the standard pattern. To my knowledge, there has been no further exploration of imagery and directed forgetting. Because of the absence of an imagery effect here, it is not clear whether we can draw any conclusion about the interplay between these two instructional manipulations. However, it is interesting that the group told to image all words and inhibit half of them could not do so. Perhaps a strong encoding is less subject to directed forgetting.

Gross et al. (1970) studied what might be described as “shared forgetting.” Subjects heard a list of 18 word pairs, with one word in each pair assigned a value of 50 times the other (25¢ vs. ½¢). Subjects took part in dyads, with one member of the dyad determining which word in each pair was to be remembered by each subject. After study, subjects were tested on either their own words or those of their partner. In essence, there were really two R versus F cues at encoding: Remember the more valuable items, and remember your own items. In Experiment 1, subjects did recall their own words (.49 high value, .39 low value) better than their partner’s words (.24 high value, .19 low value), also showing an effect of value. This was not the case in Experiment 2: Neither manipulation mattered when the test was three-alternative recognition using related distractors (all means around .46). Gross et al. concluded that the critical mechanism was differential rehearsal, which subjects used primarily when they expected a recall test.

An important study for several reasons is that of Davis and Okada (1971). Subjects studied three 64-word lists at 1 sec per word, followed by 1 sec per cue, with half R words and half F words in each list. For the first two lists, recall of F words was discouraged and recall of R items (.26) was much higher than that of F items (.03). For the third list, subjects were asked to recall both the R items (.27) and the F items (.05), making little difference. It is noteworthy that the R words showed a standard bowed serial position curve, but the function for the F words was flat (although a floor effect may have been responsible).

The main goal of Davis and Okada was to examine Yes–No recognition performance following directed forgetting. The result was clear: There was

a strong advantage for R items (.70) over F items (.50). Davis and Okada were also among the first to test retention of the cues originally associated with each word. Of the words correctly recognized, .83 of R words versus only .64 of F words were correctly labeled. So subjects had decent ability to discriminate R from F items, at least for correctly recognized words. But the key result in Davis and Okada's study was the strong remember-forget difference in recognition, which they claimed was inconsistent with any retrieval-based account (e.g., inhibition), and more in keeping with a selective rehearsal effect on encoding.

Waugh (1972, Experiment 2) used directed forgetting to investigate why long lists are harder to remember than short ones. Subjects studied normal 20-word and 40-word lists, plus 40-word lists with a midlist cue to "erase" the first 20 words. Recall accuracy was .26 in the 20-word list. For the normal 40-word list, recall accuracy was .10 for the first 20 words and .19 for the second 20 words. For the "erase" 40-word list, recall accuracy was .03 for the first 20 words and .24 for the second 20 words. Thus, there was almost no directed-forgetting cost. In her third experiment, Waugh showed that the erase condition was essentially unchanged if half of the items in the second half of the list were repetitions of items in the first (or "erased") half of the list (cf. Timmins, 1973, 1975). Her conclusion was that information had to be re-rehearsed periodically to remain accessible, and that the likelihood of such rehearsal was greater in shorter lists, including lists where subjects were permitted to forget some of the items, thereby leading to a functionally shorter list.

Timmins (1974) explored variations in processing time on the effect of an F cue. In each list, subjects saw 15 words, each set of 5 followed by a cue. "Forget set 1" could appear after 5, 10, or 15 words; "forget set 2" could appear after 10 or 15 words; otherwise, the cue was "continue." Then, just prior to testing, subjects were cued to recall sets 1 and 3 or sets 2 and 3. (Unfortunately, Timmins did not report recall of F words or recognition of R words.) Recall of set 3 was uniformly good (.80 to .86) and unaffected by the presence or location of the F cue. For set 2, recall accuracy was .41 without a set 1 F cue, and declined from .74 to .51 to .43 as the F cue for set 1 moved to later in the list. For set 1, recall was .49 without a cue to forget set 2, and declined from .61 to .47 as the F cue for set 2 moved to later in the list. Thus, F words interfered less when subjects could act on the cue right away. Indeed, placing the F cue after the list eliminated its value (see also Spector et al., 1973). These claims were supported by the recognition data for F words: Recognition improved for both set 1 items (.45, .66, .64) and set 2 items (.42, .59) as the F cue moved to later in the list, presumably because those items benefited from more processing with later F cues.

Woodward et al. (1973) showed that items are minimally rehearsed until their cue appears in the item method. What Timmins added is that when

the F cue is delayed in the list method, subjects cannot suspend processing so they rehearse as they would for R items, eroding the potential value of the F cue. Holding words in memory that will eventually be designated as F words improves memory for those words, but at the expense of the R words in the list. This pattern is entirely consistent with a rehearsal account.

Finally, taking a more indirect approach, Kausler and Settle (1974) used homophones (e.g., kernel, colonel) to investigate false alarms in recognition. Subjects studied 22 F words and 38 R words for 3 sec each using the item method. Among the 60 recognition test items were 12 studied words, 12 F-word homophones, 12 R-word homophones, and 24 unrelated words. Homophones produced more false alarms for R words (.16) than for F words (.12), relative to control items (.10), averaging over experiments. Thus, even an indirect test suggests that F items are less accessible, undermining demand characteristic concerns that subjects simply withhold F items on the test.

### **Summary of the Golden Age**

Over the first decade of directed-forgetting research, several things became clear. First, the effect of an F cue was consistently quite powerful, independent of the specific procedures used. The effect was apparent in both short-term and long-term memory. However, its consistency in recall was not reflected in recognition. Many of these early studies contrasted the case where a list contained an F cue to the case where all items were to be remembered. Thus, the comparison was between R items accompanied by other R items or by F items. Both benefits for R items and costs for F items were reported, primarily in accuracy of recall (and sometimes recognition) but also in latency.

This first epoch also introduced all of the major theoretical accounts that would come to dominate subsequent explorations of directed forgetting right up to the present. The erasure idea—that F items were culled from memory—was quite quickly discarded. Apart from the work of Weiner and his colleagues, few investigators subscribed to the notion that F items were repressed (or inhibited) at the time of retrieval. But the ideas of segregation and differential rehearsal, advocated by Bjork, and of selective search, advocated by Epstein, garnered much attention and support. The stage was set to expand the study of directed forgetting, a promising new memory tool.

### **THE SILVER AGE: 1975–1985**

Over the next decade, directed forgetting became more of a subject of study in its own right. What made directed forgetting work? How did the encodings of R and F items differ? To what situations might the manipulation be

extended? How did directed forgetting differentially influence the emerging new types of memory tests? These were a few of the questions that guided research on directed forgetting through the formative years of its “adolescence.” As in the early years, there was also one investigator who was featured in this era, so I end with an overview of his work. This work is especially relevant because it also laid the foundation for a shift in the preferred explanation of directed-forgetting effects toward inhibitory mechanisms, a shift that was occurring throughout cognitive psychology (see, e.g., Arbuthnott, 1995). To begin, though, I consider the last of the short-term studies before turning to the burgeoning long-term memory research.

### **Last of the Short-Term Memory Studies**

Most work on short-term effects of directed forgetting is tied to the Golden Age. But a few short-term memory studies using response latency trickled over into the more eclectic Silver Age.

Using recognition latency as the dependent measure, Epstein, Wilder, and Robertson (1975) tested whether the search set was smaller when a list contained F items. Subjects studied four-pair trigram-word lists with each pair presented for 4 sec. F cues appeared either before or after the list (remember only the first two or the last two pairs); R cues always followed the list. Each of the 96 lists was followed with a one-pair test to which the subject responded yes (for studied pairs) or no (for re-paired items) as quickly as possible. On “no” trials, the re-pairing could be within set or across set, but was always within list. Averaging over response types (because they behaved alike), performance was slowest and most accurate on R trials (2002 ms, .23), intermediate on F-after trials (1890 ms, .17), and fastest and least accurate on F-before trials (1606 ms, .10). Epstein et al. concluded that the presence of an F cue reduces the search set more for a prior than for a subsequent F cue. The fact that a postlist F cue nonetheless led to more rapid search suggested to them that the R and F items are separated and can be searched at least partially independently.<sup>6</sup>

Howard (1976) reported much the same pattern (Epstein et al., 1975, had modeled their study on an earlier Howard technical report). She used shape-picture pairs, presenting four in each block at a rate of 6 sec per pair, with the first two distinguished from the second two by slide color. Cues were delivered either at the middle of the block or after it. Once again averaging over similarly behaved response types, R trials were the slowest and most accurate (1773 ms, .13); however, F-after (1615 ms, .10) and F-during (1547 ms, .03) were more comparable. Presumably because subjects now

---

<sup>6</sup>A cautionary note must be sounded: The speed-accuracy tradeoff pattern evident in this study and the next one complicate interpretation.

had to encode and process the first two items at least until they received the midlist cue, which they did not have to do in the F-before condition of Epstein et al., the F-after and F-during conditions differed less. As in many directed-forgetting studies (see especially Epstein's early work), being able to avoid encoding is quite different from encoding and then trying to forget. Howard argued that the R and F items were segregated, possibly with the aid of selective rehearsal.

Virtually all directed-forgetting studies have used verbal materials: Cruse and Jones (1976) used auditory tones instead. If differential rehearsal alone caused directed forgetting, and if tones could not be rehearsed, then directed forgetting should not affect tones. They modified the Sternberg (1969) memory-scanning paradigm, following the lead of Kaminsky and DeRosa (1974), who showed that when the search set was divided into two subsets (letters and digits), subjects could restrict their search to one of the subsets. Cruse and Jones' idea was that, for a memory set of both R and F items, subjects would only have to search the R items. Using set sizes of one to four tones, with one to three of those to be remembered, they tested subjects with R items (yes) versus new items (no) as probes. They also included a few F probes, also "no" responses. Study tones were presented for 500 ms each with a 500 ms R or F cue between successive tones and a test tone after a 1200 ms gap.

When only one tone had to be remembered, response times were faster (in the range of 325 ms) than when more than one tone had to be remembered (425–450 ms). Thus, there was a cue effect, but limited to a single R item. Furthermore, there was still an effect of the F items: The search slope with a single R item, although shallow, was nonzero. Cruse and Jones argued that because they observed directed forgetting for a single R item, and assuming that rehearsal was not possible for tones, then selective rehearsal was not an adequate explanation. They preferred a selective search explanation, noting that ability to segregate the search in short-term memory might suffer as more items in the R set increased the memory load, disrupting the segregation.

### **The Growth of Long-Term Memory Studies**

Most of the studies after the first 10 years involved longer lists. In part, this was no doubt due to the increasing impact of the levels of processing framework ( Craik & Lockhart, 1972), with its emphasis on the processing and retention of larger amounts of material.

Rakover (1975) manipulated test expectations. Subjects studied a list of nine trigrams and nine 3-digit numbers, with one group told to remember trigrams only, and the other told to remember numbers only. This cue was given prior to study, nonoptimal because subjects can simply ignore the F

set, instead of encoding and then having to forget them. The list was shown 1, 3, 5, 7, or 14 times. After study, all subjects were required to free recall the trigrams only. For the digits group, accuracy was only about .22 overall, and was relatively unaffected by number of study trials: .14, .19, .24, .30, and .20. However, for the trigram group, recall was much better overall (.73) and rose with number of presentations: .44, .71, .64, .87, and .98. Rakover concluded that F items were dropped from short-term memory whereas R items continued to be processed, essentially a selective rehearsal account.

In a follow-up study, Rakover (1976) manipulated list length (3 vs. 12 of each type of item). Lists were made up either of digits and trigrams, only digits, or only trigrams, and subjects were told in advance which to remember. When tested, all subjects had to recall trigrams only. For lists containing both types of items, trigram recall was much better for the R group (.91 and .50 for 3 and 12 item lists, respectively) than for the F group (.11 and .05, respectively). For lists containing only one type of item, the pattern was similar: for the R group, .83 and .47; for the F group, .29 and .14. Once again, Rakover argued that F items received less postpresentation rehearsal.

Rakover and Kaminer (1978) tested the rehearsal idea in one more study. Subjects studied 15 trigrams at a rate of one every 2 sec. In one condition, they saw each trigram once and were to remember all of them. In three other conditions, they saw each trigram twice, with the lag between repetitions manipulated. One condition required remembering all of the trigrams, one required remembering any underlined trigram, and one required forgetting any underlined trigram (always on its second presentation). Recall accuracy for F items rose from about .08 to .21 with lag in the group told to forget underlined trigrams. Recall accuracy was a steady .20 to .30 for the three other groups. Recognition showed a similar lag pattern. Rakover and Kaminer suggested that the increase with lag in the F condition was due to F items being increasingly more rehearsed with longer gaps between successive presentations. With a long enough gap, there was almost no cost for F items.

Jongeward, Woodward, and Bjork (1975) promoted the idea of segregation. Subjects saw five 32-word lists divided into 8 four-word blocks, with each block followed by a cue that either preceded or succeeded a 3-sec rehearsal period. That cue required remembering all four words in the block, only the first two, only the last two, or none of them. After all five lists had been studied and tested, there was a final recall test for all words from all lists. For the rehearsal group, recall of R words from lists containing F cues (.38 immediate; .18 final) was better than recall of the corresponding words from lists of only R words (.30 immediate; .13 final); for the no rehearsal group, these two conditions did not differ (F cues: .36 and .16; R cues: .34 and .15). Both groups recalled F items poorly, at less than .10. Selective rehearsal benefited R items, even given identical search set sizes

for R and F items. Jongeward et al. were especially impressed at the ability of subjects to keep the R and F items apart, arguing that segregation was the critical mechanism.

In contrast, MacLeod (1975) argued for a selective rehearsal interpretation based on the effects of F cues over a long retention interval, in this case 1 or 2 weeks. In Experiment 1, subjects studied 60 words made up of 3 each from 20 semantic categories. The setup was 3 sec for the category name, 3 sec each for the words, and 6 sec for the cue, with half of the categories being R cued and half F cued. Either immediately after study or 1 week later, subjects completed category-cued recall and three-alternative recognition tests. The R-F difference was the same regardless of test type or retention interval: There was a constant .10 advantage for R over F items on both tests at both retention intervals. Assuming that set-differentiating information would be lost before item-specific information, then a selective search account predicts that the R-F difference should diminish with retention interval; it does not. Moreover, the effect is constant in recall and recognition. Both results are inconsistent with selective search, but quite in keeping with selective rehearsal.

In Experiment 2, MacLeod switched to unrelated items. Subjects saw each of the 38 studied words for 3 sec, followed by its cue for 3 sec. A three-alternative recognition test was administered 1 or 2 weeks later. For each item, subjects either were told the original cue associated with that item or they had to recall that item's original cue. There was a strong R-F difference, and recognition declined with retention interval (R-1 = .68; F-1 = .53; R-2 = .64; F-2 = .47). Providing subjects with the original cues did not matter, conflicting with the selective search account. Overall, MacLeod favored a selective rehearsal explanation because of the persistence of the effects over long retention intervals.

To examine rehearsal more closely, Wetzel and colleagues manipulated the timing of F cues. In the first of this series, Wetzel (1975) combined a levels-of-processing manipulation of encoding with directed forgetting. Subjects heard three lists of 16 words each under one of two encoding conditions:<sup>7</sup> They were either to write down the word, or to write down an adjective to modify the word. In Experiment 1, each word was presented for 5 sec and then an R or F cue was presented, with 1 sec before the next word appeared. In Experiment 2, this timing was reversed. Cues were randomized through the list, with half of the words receiving each cue. Subjects recalled each list immediately, and then had an unmentioned final recall and a final Yes-No recognition test after all of the lists.

---

<sup>7</sup>There were actually three conditions, the third being a rhyme encoding, but because the rhyme-encoding and adjective-encoding conditions did not differ, and because the rhyme-encoding condition was not included in Experiment 2, I deleted it.

The patterns over experiments were virtually identical, so I only report Experiment 1. On immediate recall, there was a large R-F difference for both rehearse words (R: .61, F: .13) and adjective words (R: .36, F: .12). Serial position analyses showed recency for both R and F words but primacy for R words only. Most intriguing, the levels-of-processing manipulation affected only R words. Final recall showed a similar, though muted, pattern (rehearse, R: .33, F: .13; adjective, R: .21, F: .12). Final recognition showed an even more muted pattern (rehearse, R: .75, F: .64; adjective, R: .67, F: .59). Wetzel argued that the potential of differential processing prior to the cue was only realized for R words: The F cue stopped further processing of F words. This was entirely consistent with a selective rehearsal account, with items held in limbo until the instruction tells the subject what to do, so that only R items benefit from prior processing. The fact that the duration of the cue interval had little impact is also consistent with subjects waiting for the instruction.

To delve into cue delay, Wetzel and Hunt (1977, Experiment 2) again manipulated the time between successive words. For half of the subjects, this was done within list. In the long-delay group, subjects had 1, 4, 8, or 12 sec before the cue and 1 sec after it; in the short-delay group, they had 1 sec before the cue and 1, 4, 8, or 12 sec after it, thereby keeping the total interval between words comparable. For the other half of the subjects, the rehearsal intervals were the same, but were manipulated between lists, such that a given list had only one of the four intervals. Subjects heard four 32-word lists, with a random half of the words in each list followed by each cue.

In the within-list case, there was little effect of the four rehearsal intervals, with short-delay R words (.64) showing better immediate recall than long-delay R words (.46) regardless of delay. In the between-list case, performance rose steadily as a function of rehearsal interval: Short-delay R words (.36, .57, .75, .84) again were better recalled than long-delay R words (.38, .47, .60, .61). F words showed the opposite pattern: Long-delay words (between list: .07, .10, .13, .14; within list: .17, .10, .11, .18) showed some benefit of rehearsal compared to short-delay words (all around .04). Presumably, the short delay helped R items by reducing interference before further processing; the long delay helped F items by permitting a little rehearsal prior to suspension of processing. Final recall showed similar although attenuated effects. Recognition also showed a strong R advantage, but all of the long-delay performance shifted upward, demonstrating the important role of maintenance rehearsal in recognition. Also, recognition of F words was relatively better than was recall of F words, although still far from the level of R-word recognition. Wetzel and Hunt saw these results as strong evidence of selective rehearsal in directed forgetting. The contrasting impact of delay on the R and F items is a particularly nice illustration of the importance of duration versus kind of rehearsal.

Tzeng, Lee, and Wetzel (1979) looked at how well subjects could judge when R versus F items had appeared in a list. In Experiment 1, subjects studied a 40-word list at 2 sec per word and 3 sec per cue, with half of the words randomly assigned R and F cues. Recall was much better for the R words (.65) than for the F words (.17); recognition showed a less dramatic but still reliable difference (R: .95; F: .87), perhaps limited by a ceiling effect. But the focus was on the last test, where subjects judged the serial position of each word they had recognized in terms of five-item successive blocks of the list. Position judgments were very accurate for R words, but extremely poor for F words.

In Experiment 2, Tzeng et al. used another measure of temporal judgment. On the same list, recall of R words (.61) was again better than recall of F words (.22). Subjects then judged for pairs of words which one had appeared earlier or later in the list. Accuracy declined from .83 for R-R pairs to .71 for R-F pairs to .60 for F-R and F-F pairs, where the first member of each pair was the one that actually appeared first in the list. Tzeng et al. argued that accessibility of the first item at the time of encoding the second one is crucial for the pairwise judgment, explaining the advantage of an R item being first in the pair. Overall, Tzeng et al. saw their studies as evidence that temporal coding occurs quite automatically for R items but not for F items. Temporal information could even be part of the basis for the segregation of the R- and F-item sets that must underlie selective rehearsal (cf. Bjork, 1972).

Deviating from strict chronological order, a later study by Jackson and Michon (1984) is relevant here. First, they nicely replicated the Tzeng et al. (1979, Experiment 1) order results using both concrete and abstract words. Then, based on an examination of the correlations between judged position and actual position, they went on to argue that unrecalled R items, like F items, might not have been entered into the rehearsal set, which apparently is crucial for good retention of temporal order, and therefore usually favors R words. In arguing that temporal order information is not automatically encoded, they made it clear that there is very little temporal information for F items, consistent with F items not being rehearsed at the time of encoding.

Like Wetzel (1975), Horton and Petruk (1980) focused on the type of encoding that subjects engaged in and the timing of the cue. Each word in the three 24-word lists was presented for 2 sec followed by a 2-sec cue. A further 5 sec was assigned to encoding task—structural (write a word starting with the same letter), phonemic (write a rhyming word), or semantic (write a word from the same category). For half of the subjects, cues preceded encoding tasks; for half of the subjects, cues followed encoding tasks. For one third of each of these groups, the words were unrelated; for one third, the words were categorized and the cues honored the categories (e.g., forget all fruits, remember all animals), and for one third, the lists were

categorized but the cues divided categories (e.g., forget half of the fruits and half of the animals, remember the other half of each).

In immediate recall, R words (.51) were better remembered than F words (.23). As well, unlike Wetzel (1975), there was a typical levels-of-processing effect, with semantic (R: .50; F: .31) better than phonemic (R: .35; F: .20) and structural (R: .36; F: .20). Note that the R advantage was greatest in the semantic condition. Collapsing over encoding condition, unrelated lists were more poorly recalled (R: .28; F: .13) than lists of honored categories (R: .44; F: .29) or divided categories (R: .49; F: .29). As for cue location, R items benefited from prior cues (.54) versus subsequent cues (.48), but this manipulation did not matter for F items (.22 and .23, respectively). Final recall data patterns were quite similar. Recognition also showed a reliable effect of instructional cue.

Horton and Petruk emphasized the relation between set differentiation and level of processing. The R and F sets were better separated, allowing a greater advantage for R items, when the cues preceded the words and when the words were processed more semantically. Under selective rehearsal, more semantic processing should have improved memory for both R and F items, undermining segregation. In fact, the opposite happened. Horton and Petruk argued that R and F items were stored separately, casting their vote for selective search (cf. Epstein, 1972).

At the outset, I mentioned Slamecka's (1968) part-list cuing effect as related to directed forgetting. When shown part of the studied list at the time of testing, subjects recall the remainder of the list more poorly than without such cues. Roediger and Tulving (1979) bridged these two topics. In Experiment 1, subjects studied two 64-word lists made up of eight items from each of eight categories, with the words from every category beginning with the same eight letters. The lists were presented at a 2-sec rate, one category at a time. Following study, subjects received either a standard free recall test or one of two "exclusion" tests on both lists. In one exclusion test, subjects were told not to recall items from four specified categories; in the other, they were told not to recall items beginning with four specified letters. The two lists were studied and tested one at a time.

In contrast to the usual pattern in directed forgetting, recall was slightly poorer when subjects had to exclude some items (.29) than when they were free to recall any items at all (.33). Switching to excluding two or six categories or letters in Experiment 2 hardly changed the pattern. In Experiment 3, subjects studied either a 30-word list (six categories) or a 60-word list (12 categories). They then recalled all of the words or, in the 60-word list, only a specified half of the categories corresponding to the 30-word list. Not surprisingly, more words were recalled from the 60-word list, about .53. But recall of the 30-word list was about .62 whereas recall of only a specified 30 of the words in the 60-word list was .47. Exclusion of part of the list resulted

in poorer recall than was observed for a corresponding shorter list, analogous to their earlier part-list cuing experiments.

In directed forgetting, the cue is given during study. In part-list cuing, the cue is given at the time of retrieval. Roediger and Tulving's main claim was that a cue to forget will ordinarily be successful only when given during encoding. If given at the time of testing, the cue would have to uniquely specify what was to be recalled (e.g., one of only two categories) in such a way as to permit rapid recall, thereby benefiting by avoiding output interference.

In a more naturalistic study related to the upcoming Geiselman studies of text recall, Golding and Keenan (1985) gave subjects a set of verbal directions to remember. One group received the basic directions, a second received the basic directions plus an extra item, and the third received the basic directions, the extra item, and an instruction to forget the extra item. On an immediate recall test, performance on R information was worse for the two groups with the extra item (R: .30; F: .28) than for the group with just the basic directions (.35), but there was no R-F difference. The same pattern occurred on a verbal recognition test (R: .58; F: .64; basic: .72). Furthermore, neither recall nor recognition showed any difference in performance on the F items. It was only when subjects drew the map that the F cue helped memory for the R items by reducing the memory load (R: .53; F: .74; basic: .69). They concluded that subjects encode F information together with R information and mark the F information as wrong, to help keep track of what is in fact correct.

### **Geiselman and Colleagues: From Selective Rehearsal to Repression**

Edward Geiselman was the most active researcher in the area of directed forgetting during this middle era. It is probably not surprising that he was a colleague of Bjork's at UCLA during much of this time. Initially, his explanation of directed forgetting relied primarily on selective rehearsal, but his perspective changed quite dramatically over the 10-year period.

Geiselman's early work used sentences. In the first study (Geiselman, 1974), subjects studied 10-sentence passages, with each sentence shown for 5 sec, followed by a 1-sec R or F cue and a 5-sec blank interval. The first two passages were followed by recall tests for R items only, and the last was followed by either a recall test or a sentence completion test of both R and F items. R items from the last passage were better recalled (.74) than the corresponding R items from the first two passages (.57), a directed-forgetting benefit of .17. Furthermore, recall of F items from the last passage (.40) was worse than recall of the corresponding R items, a directed-forgetting cost of .17. On the sentence completion test, R items were completed reliably better (.87) than F items (.75).

Geiselman took the subjective reports of rehearsal during the blank intervals and the symmetry of the cost and benefit results as strongly supporting selective rehearsal. He discounted an explanation in terms of earlier output of R items by replicating the experiment and requiring output of the F items first: The results were unchanged. He also noted that the size of the effect seemed smaller for meaningful materials like sentences than for individual, isolated words. This may have been the result of remembered R sentences serving as retrieval cues for related F items, thereby undermining successful directed forgetting (see Golding, Long, & MacLeod, 1994).

Next, Geiselman (1975) showed subjects 10 sentences from each of two different themes, in random order. In advance, half of the subjects were told to remember all of the sentences from both themes and half were told that they could forget all of the sentences from one of the themes. Sentences within a theme were either ordered or scrambled, with each shown for 7 sec, followed by a 7-sec blank interval. R sentences enjoyed a strong benefit in recall for lists containing F sentences (.50) relative to lists of only R items (.23). Moreover, presenting either the R or the F sentences in order benefited recall of R items. Experiment 2 used eye movements to reveal that easier encoding of the sentences and easier grouping of the themes both contributed to the advantage due to ordered presentation of the sentences. Geiselman concluded that his results supported the need for both the grouping and the selective rehearsal elements of Bjork's (1972) account of directed forgetting.

Geiselman and Riehle (1975) extended this study to show that, like R sentences, recall of F sentences benefited from being presented in order. For just the R sentences in lists containing both R and F sentences from Geiselman's (1975) study, accuracy was .64 when both R and F sentences were ordered versus .36 when both were scrambled. Having just the F sentences ordered (.53) led to slightly better recall than having just the R sentences ordered (.46). Geiselman and Riehle replicated this study, adding results for the F-sentence recall. Recall of F sentences was much better when only the F sentences were ordered (.41) than for the other conditions (all .29 or less). They concluded that ordering the sentences made segregating the two sets of sentences easier.

In the final study in the sentence-memory set, Geiselman (1977) again presented sentences from two themes in random order, this time auditorily. The two themes were hard to understand without a title. Half of the subjects were told the titles and half were not. Order within theme was again either logical or scrambled, and the benefits of order from the prior studies replicated. But for the order effects to appear, it was necessary for the theme of the F sentences to be explicit. Presumably, this permitted the sorting necessary to group the R and F sets. During the study phase, Geiselman also measured time to decide whether a sentence was from the F or the R

theme, and found that R recall was better for shorter decisions whereas F recall was better for longer decisions. This suggests that trying to remember benefits from easier comprehension, whereas trying to forget suffers from lengthened study time. Finally, Geiselman found that recognition of individual words from the sentences was better on R sentences than on F sentences only if subjects had been told the theme for the F sentences. Once again, organization was critical to permit sorting by cue.

Geiselman's sentence experiments are instructive in a couple of regards. First, directed forgetting works with meaningfully related information. However, there is the strong suggestion that relatedness among R and F items undermines the success of an F cue. Second, organization of the materials—both of the R and the F materials—helps in successfully using the cue to forget some items and remember others. Geiselman argued that the better grouping of the items under the two cue conditions permitted operations to be applied to the conditions selectively. Distinctiveness appeared to be crucial; even organizing the F items made the R items better remembered.

By the mid 1980s, Geiselman was arguing for a very different interpretation of the directed-forgetting phenomenon, one that emphasized processes operating at the time of retrieval more than those at the time of encoding, thereby restoring interest in the repression account offered early on by Weiner (1968; Weiner & Reed, 1969). Geiselman came to this new perspective through exploration of the link between directed forgetting and posthypnotic amnesia. I only consider the standard memory research here, however; studies relating hypnosis and directed forgetting are discussed in the next chapter by Golding and Long.

The critical study was that of Bjork and Geiselman (1978), in which they aimed to eliminate rehearsal. Their subjects studied two lists of 16 pairs of words for 3 sec each, with 3–12 sec of arithmetic distraction after each pair. The cue followed the distraction interval. One list was followed by immediate recall of R words; the other was followed by distraction only. Then subjects did a final recall of all items, both F and R words, followed by a final Yes–No recognition test.

The immediate recall data showed the usual large advantage for R words (.30) over F words (.06). Unlike the flat performance for F words across the 3–12 sec distraction interval, recall of R words increased from about .20 to .40, in contrast to Woodward et al. (1973). This suggests that differential rehearsal occurred despite efforts to prevent it. Final recall showed a very similar pattern, with prior testing helping R items but not F items, although interpretation is complicated by the apparent floor effect for F items on both recall tests. Final recognition was unaffected by either distraction duration or prior testing, but did show an advantage for R items (.68) over F items (.56).

Bjork and Geiselman argued that only R items are retrieved when their cue appears. If so, then forcing retrieval of F items at the time of their cue

should attenuate the effect. In Experiment 2, they required subjects to retrieve every pair upon presentation of its cue. Within-list retrieval accuracy was almost identical for both the R and F items. Immediate recall showed a substantially reduced, although still reliable, advantage for R words (.15) over F words (.11). Once again, the pattern was similar in final recall, but now the recognition difference vanished (.64 for both R and F). Bjork and Geiselman argued that their balancing of retrieval disrupted the normal segregation of R and F items, and shifted their explanation from selective rehearsal to selective retrieval.

Geiselman, Bjork, and Fishman (1983a; see also Geiselman et al., 1983b) provided converging support for this differential retrieval hypothesis. Subjects heard a list of 48 words, one every 7 sec. They were told to learn one word and to judge the pleasantness of the next one, alternating intentional and incidental learning. Halfway through the list, half of the subjects were told to forget the preceding words; the rest were told to continue remembering all of the "learn" words. Recall of the postcue R words was greater in lists with F cues (learn: .72; judge: .42) than in lists without F cues (learn: .57; judge: .30). The reciprocal pattern occurred for precue words: F-cued words (learn: .57; judge: .32) were recalled worse than R-cued words (learn: .74; judge: .47). Indeed, the tradeoff was close to perfect. But in Yes-No recognition, there was no effect of either cue or initial study mode. Geiselman et al. (1983a) argued that although selective rehearsal could explain the difference in the "learn" items, where subjects might reasonably have rehearsed R items more than F items, it was not clear how this view fit judged items. Incidentally learned items should not have attracted any rehearsal, let alone differential rehearsal. Why were judged words affected by the F cue?

One possibility was that somehow the judged and learned words became linked during study. In Experiment 2, the two sets of words were chosen from distinct semantic categories to prevent such linkage, yet the pattern in Experiment 1 persisted. Furthermore, in both experiments, subjects had difficulty identifying the half of the list in which F-cued words were presented. Experiments 3 and 4 evaluated whether subjects tended to recall the second half first for lists containing an F cue, generating output interference for items from the first half (the F items). When subjects were instructed to recall the first half before the second or the second half before the first, it made no difference at all to the other effects, discounting the output interference hypothesis.

Geiselman et al. (1983a) maintained that selective rehearsal could not explain the R-F difference in the judged items, where rehearsal was irrelevant and did not seem to occur. Instead, they argued for an inhibitory process preventing F item retrieval. This retrieval-blocking idea was consistent with three facts: (a) There was no effect of cuing on recognition, (b)

subjects could not identify list half for items prior to an F cue, and (c) the correlation between input and output order was very low for items prior to the F cue. They suggested that inhibition made sense: F items would usually not be retrieved, but could be revived if needed, perhaps by re-presentation. Geiselman et al. (1983a) did not note that their study used the list method whereas many prior studies had used the item method to deliver the F cue. This critical connection was made later by Bjork (1989) and by Basden, Basden, and Gargano (1993).

The linchpin study to support retrieval inhibition is that of Geiselman and Bagheri (1985). Surprisingly, they used the item procedure to deliver the F cue, now ordinarily associated with the rehearsal explanation. In Experiment 1, subjects saw 36 words for 5 sec each, each followed by a cue for 5 sec. Subjects recalled many more of the R words (.58) than of the F words (.10). They then studied the same list again (actually four each of the R and F items were changed), but this time all words were to be remembered. Recall rose much more for the formerly F words (.49) than for the formerly R words (.65). Most important, words not recalled on the first test were much more likely to be recalled on the second test if they had initially been F items (.45) rather than R items (.28). Their explanation was that unrecalled F words benefited from a release from retrieval inhibition in addition to the extra rehearsal that both R and F items attracted. The possibility that unrecalled R words were simply harder than unrecalled F words was discounted by showing that yoked subjects recalled these two sets of items equally well after a single study trial.

It could still be argued, however, that this does not address item selection at the *individual* level. In Experiment 2, they included R and F words, plus words judged for pleasantness. The retrieval inhibition view predicts that only the F items can benefit on the second recall from release from inhibition. On the first recall, F words (.07) were poorly recalled relative to judged words (.32) or R words (.46). On the second recall, following a study trial where all words were repeated and were to be remembered, the F words (.50) rose considerably more than either the judged words (.55) or the R words (.66). Furthermore, the conditional probability of recalling a previously unrecalled word was greatest for formerly F words (.54) as compared to formerly judged (.34) or R (.37) words.

Could unrecalled F words from the first trial have been selectively rehearsed on the second trial? In Experiment 3, the second list contained only formerly F items or formerly R items, with all to be remembered. The initial R-F difference (.44 vs. .13) almost disappeared on the second test (.60 vs. .54) and again, unrecalled former F items improved more (.51) than did unrecalled former R items (.34). Selective rehearsal of the formerly F items during the second study sequence seemed unlikely, so the evidence fit a retrieval inhibition account better. But the fact that there were more initially

unrecalled F than R items might have helped formerly F items on the second study trial. To test this idea, Experiment 4 replicated Experiment 3 but with only 10 items to be learned. A faster presentation rate (3 sec/word) was used to minimize rehearsal opportunities and to lower overall performance. The initial advantage for R (.29) over F (.05) items disappeared on the second recall, with formerly F words (.70) now better learned than formerly R words (.56).

Geiselman and Bagheri (1985, p. 61) argued that, in addition to encoding factors, "a strong case can also be made for a significant role of memory retrieval inhibition." Presenting unrecalled F and R items a second time as R items helped the F items more. This was taken as evidence that F items, unlike R items, benefited additionally from release of retrieval inhibition on a second study. Whether item selection on the second study trial was completely ruled out is uncertain, but the results have proven to be tantalizing with regard to possible inhibition of F words.

Geiselman and Panting (1985) suggested four possibilities for why repetition of F items from list 1 as R items in list 2 led to equivalent performance for formerly R and formerly F items: (a) Output interference affected F items in list 1; (b) more effort was required on formerly F items on list 2; (c) there was intentional withholding of F items on list 1; and (d) item selection favored formerly F items on list 2. To evaluate these possibilities, in Experiment 1 subjects saw six categories of eight words each, each word presented for 3 sec and each category followed by a 3-sec cue. Subjects then attempted to recall all of the items from one category at a time, with category cues provided by the experimenter. In this way, either all of the F or all of the R categories could be tested first. Following recall, subjects saw all of the R or all of the F words again, and were instructed to remember all of them.

The effect was quite clear on the first test, whether F categories (R: .56; F: .47) or R categories (R: .66; F: .45) were tested first. However, the difference vanished on list 2 (all conditions around .77). As previously, list 2 recall conditional on list 1 nonrecall was greater for formerly F items (.59) than for formerly R items (.34). The basic pattern—a difference on list 1 but not on list 2—was taken as consistent with the retrieval inhibition view, although the larger R-F difference with R recall first did indicate a contribution of output interference. Because a yoked group of subjects, who learned the unrecalled items from list 1, showed no R-F difference, item selection was considered unlikely, although a yoking design does not rule out the possibility of subject-specific item selection. Because only either R or F items were relearned, differential effort was seen as unlikely; because F items and R items were separately cued, suppression of F items was seen as unlikely. Of course, although plausible, these arguments are speculative: Subjects could still have operated strategically despite these design features.

Finally, a study by Geiselman, Rabow, Wachtel, and MacKinnon (1985) explored strategic processing in directed forgetting at the encoding, forget-