

DISSERTATION

A MULTIPLE CUE HYPOTHESIS FOR THE TESTING EFFECT

Submitted by

Shana K. Carpenter

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2004

UMI Number: 3160086

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3160086

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

November 5, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY SHANA CARPENTER ENTITLED A MULTIPLE CUE HYPOTHESIS FOR THE TESTING EFFECT BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Benjamin C. ...

Charles ...

Frank J. ...

Edward ...

Advisor

Edward ...
Department Head/Director

ABSTRACT OF DISSERTATION

A MULTIPLE-CUE HYPOTHESIS FOR THE TESTING EFFECT

Three experiments investigated the role of variable processing in the testing effect. All three experiments utilized a paired-associate paradigm in which the same target items were presented, followed by an intervening cued-recall test or additional study opportunity, followed by a final free-recall test over all target items. In Experiment 1, related cue-target pairs (*dough: bread*) were presented, followed by either an intervening test in which the cue was either the same or different from the one encountered during initial presentation (*dough: _____, toast: _____*, respectively), or by an intervening study opportunity in which the cue was either the same or different from the one encountered during initial presentation (*dough: bread, toast: bread*, respectively). In Experiment 2, cue-target pairs were the same from initial presentation to intervening test or study, and were either highly related (*dough: bread*) or weakly related (*basket: bread*). In Experiment 3, unrelated cue-target pairs (*basket: bread*) were presented, followed by either a yes/no recognition test or additional study opportunity in which the correct target was presented after either 0 incorrect targets (*basket: bread*), 1 incorrect target (*basket: wicker, basket: bread*), 2 incorrect targets (*basket: wicker, basket: eggs, basket: bread*), or 3 incorrect targets (*basket: wicker, basket: eggs, basket: carry, basket: bread*).

A significant testing effect was obtained in all three experiments. In addition, a significant advantage in retention was obtained for target items presented with different as opposed to same cues (Experiment 1), a significant Trial x Cue Interaction was

obtained in which retention for tested items benefited from unrelated as opposed to related cues whereas retention for studied items was unaffected by cue-target relatedness (Experiment 2), and finally, retention for studied items but not tested items benefited significantly from a greater number of incorrect targets such that the relative advantage for tested over studied items decreased with each additional incorrect target (Experiment 3). Results of these three experiments suggest that retention for both tested and studied items can benefit from variable processing, and that the advantage for tested over studied items is greatest under conditions that promote the sampling of extra information during retrieval.

Shana K. Carpenter
Department of Psychology
Colorado State University
Fort Collins, CO 80523
Fall, 2004

TABLE OF CONTENTS

Title Page	i
Signature Page	ii
Abstract	iii
Table of Contents	v
Chapter 1 – Introduction	1
Chapter 2 – Variations on the Concept of Variability	11
Chapter 3 – Outline of Experiments	20
Chapter 4 – Experiment 1	28
Chapter 5 – Experiment 2	42
Chapter 6 – Experiment 3	48
Chapter 7 – General Discussion	61
References	76
Appendix 1	84
Appendix 2	86
Appendix 3	88
Appendix 4	90
Appendix 5	92
Appendix 6	94

CHAPTER I

INTRODUCTION

Tests as Memory Enhancement Opportunities

Although a memory test has traditionally been conceptualized as a measurement of the contents of memory, it is by now abundantly clear that a test also *enhances* memory for retrieved information. Memory for previously-tested information is often superior to memory for non-tested information, even in the absence of feedback or additional study (Allen, Mahler, & Estes, 1969; Carpenter & DeLosh, in press; Cull, 2000; Darley & Murdock, 1971; Landauer & Eldridge, 1967; Nungester & Duchastel, 1982; Petros & Hoving, 1980; Postman & Phillips, 1961; Slamecka & Katsaiti, 1988; Wheeler & Roediger, 1992). Such enhancement in memory as a result of testing has been commonly referred to as the *testing effect*.

Past research on the testing effect has shown that it cannot be accounted for by the duration of exposure to tested material. Although retrieval brings an item to mind and may therefore be akin to an additional presentation of the item, the testing effect is still obtained in experimental designs that directly compare memory for tested items to memory for items that are re-presented but not tested (Allen et al., 1969; Carrier & Pashler, 1992; Kuo & Hirshman, 1996, 1997). It appears, then, that there is something unique about the act of retrieval, beyond mere exposure to the material, that underlies enhanced retention due to testing.

The specific retrieval mechanisms that account for enhanced retention are currently unknown. However, there are certain conditions of retrieval that seem to influence the strength of the testing effect in very consistent ways. Specifically, each of the following represent manipulations that are designed to decrease an item's accessibility at the time of an intervening test, but actually lead to better long-term retention at the time of a later memory test.

1. *The lag mediation effect.* The testing effect is stronger if a lag, or delay, is inserted between the presentation of to-be-learned material and the intervening memory test (Bjork, 1988; Landauer & Eldridge, 1967; Rothkopf & Coke, 1963, 1966; Whitten & Bjork, 1977), or if a lag is inserted between the intervening memory test and final memory test (Runquist, 1983, 1986; Wenger, Thompson, & Bartling, 1980).

2. *The test type mediation effect.* The testing effect is stronger when a recall as opposed to a recognition intervening test is administered (Bartlett, 1977; Bjork & Whitten, 1974; Carpenter & DeLosh, accepted pending revisions; Glover, 1989; McDaniel & Masson, 1985).

3. *The interference mediation effect.* The testing effect is stronger when the intervening test conditions promote interference. For example, Cuddy and Jacoby (1982; see also Gotz & Jacoby, 1978) manipulated type of intervening distracter material (similar vs. dissimilar to the target information) before the intervening test, and found that retention was better for those items that were given similar distracter material during the retention interval. Similarly, Whitten and Leonard (1980) found that an increased number of related distracters on an intervening recognition test improved later retention.

Each of these three mediating effects demonstrates that retrieval conditions rendering information *less* accessible during the time of an intervening test—even if they promote difficulty and increased errors—render that information *more* accessible during the time of a final test. A viable theoretical explanation of the testing effect must be able to account for the pattern of results associated with all three of these mediating variables. The most well-known theoretical explanations for the testing effect are reviewed below, along with the capability of each explanation to account for the lag, test type, and interference mediation effects.

Theoretical Explanations of the Testing Effect

The Total Time Hypothesis

The earliest known explanation for the testing effect is based on the notion that testing results in a re-presentation of an item via the act of retrieval. Tested items, therefore, are exposed twice as many times as non-tested items, resulting in the possibility that the testing effect may be an artifact of presentation time. Such reasoning was consistent with the already existing *total time hypothesis* (Bugelski, 1962; Cooper & Pantle, 1967), which states that the amount of information to be learned depends on the amount of time spent learning the material, regardless of the way in which the total learning time is distributed across trials.

Although some studies seemed to support this hypothesis by demonstrating that test trials did not benefit memory beyond experimenter-provided re-presentations of the material under conditions of immediate testing (Birnbaum & Eichner, 1971; Bregman & Weiner, 1970; Donaldson, 1971; Tulving, 1967), it is clear by now that the total time hypothesis cannot account for the pattern of results associated with the testing effect. The

hypothesis fails to explain why test trials benefit memory beyond experimenter-provided re-presentations of the material under conditions of non-immediate testing (see Carrier & Pashler, 1992; Kuo & Hirshman, 1996, 1997). Furthermore, the total time hypothesis cannot adequately explain the lag, test type, and interference mediation effects under conditions of equal time spent on test and study trials.

Transfer-Appropriate Processing

The principle of transfer-appropriate processing (TAP) states that memory benefits to the extent that the operations required by the final memory test match those used in the initial learning of the material (Morris, Bransford, & Franks, 1977). The testing effect may be explained from a TAP perspective by proposing that a higher degree of similarity exists in the processing conditions between an intervening test and final test, relative to the processing conditions between an intervening no-test condition (or study opportunity) and final test.

McDaniel and Fisher (1991) compared participants' memory for general knowledge facts after an intervening test versus additional study opportunity. The final memory test consisted of questions that were phrased in a way that was either similar or different from the way in which they were phrased on the intervening test. A comparison of the rate of final recall of the items successfully retrieved on the intervening test showed that final retention was better when the test questions were phrased similarly than when they were phrased differently.

Similarly, McDaniel, Kowitz, and Dunay (1989; see also McDaniel & Masson, 1985) used a cued-recall test comprised of either semantic or phonemic cues, and arranged the conditions so that the cues provided on the intervening and final tests were

either compatible (phonemic-phonemic or semantic-semantic) or incompatible (phonemic-semantic or semantic-phonemic). McDaniel et al. found that final cued-recall performance was best when the type of cue provided on the final test matched the type of cue provided on the intervening test. Final cued-recall performance was not enhanced, relative to the no-test control group, when the cues given on the intervening and final test were not of the same type.

However, there is also evidence that the match between intervening and final test conditions may not provide a complete account of the testing effect. McDaniel et al. (1989) observed a trend showing that semantic cues on the intervening test produced better retention on the final test regardless of the type of cue given on the final test. Likewise, a separate study by McDaniel and Masson (1985) reported a main effect for type of cue provided on the intervening test, such that final test performance was most enhanced by an intervening test that utilized a semantic cue.

Even more problematic for the transfer-appropriate processing hypothesis are the results of a study by Glover (1989), who examined participants' memory for ideas from a passage by using a recognition, cued-recall, or free-recall test, then subsequently gave a final test that was also based on recognition, cued-recall, or free-recall. The types of intervening and final tests were combined factorially such that some participants received compatible intervening and final tests, whereas others received incompatible intervening and final tests. Glover found that a free-recall intervening test led to the best retention, followed by cued-recall, then recognition, and this pattern held regardless of the type of final test. Thus, the transfer-appropriate processing hypothesis does not seem to fully

explain the test type mediation effect, because recall intervening tests would be more beneficial than recognition intervening tests only if the final test is also based on recall.

The transfer-appropriate processing hypothesis could explain the lag mediation effect if the strength of the testing effect increased when the lag between presentation and intervening test(s) was more similar to the lag between intervening test(s) and final test. In support of this, some studies have shown that longer intervening test lags are superior to shorter intervening test lags when the final test lag is also long (Landauer & Bjork, 1978). However, other data suggest that memory retention benefits more from longer as opposed to shorter intervening test lags, even when the final test lag is short (Cull, 2000; Cull, Shaughnessey, & Zechmeister, 1996). The fact that longer as opposed to shorter intervening test lags benefit retention, regardless of the length of the final test retention interval, suggests that the match in lag conditions between intervening and final tests does not provide a complete account of the lag mediation effect.

Finally, the transfer-appropriate processing hypothesis does not seem to adequately explain the interference mediation effect, especially when final test conditions do not contain the same type of interference as the intervening test conditions. For example, Cuddy and Jacoby (1982) administered a cued-recall intervening test over paired-associate verbal items, and this test occurred after participants were presented with several related (similar verbal paired-associates) distracter items, or several unrelated (math problems) distracter items. On a final cued-recall test over all items, memory retention was better for items that were tested under interfering conditions (verbal paired-associate distracters) as opposed to non-interfering conditions (math problem distracters). The potential for interference during the final test was reduced by providing the cued-

recall test free from distracter material. Despite this, however, conditions that *increased* interference during the intervening test led to better performance on the final test. Thus, it seems that an explanation for the testing effect based on the match in conditions between intervening and final tests does not provide a complete account of the mediation effects described earlier.

Elaborative Processing

The elaborative processing hypothesis proposes that the testing effect is not a result of the match in processing conditions between an intervening and final test, but rather is a result of the type of processing at the time of the intervening test. Specifically, the view proposes that because items are not readily accessible on test trials, more elaborative processing is necessary to recover them from memory, and such elaborative processing is what accounts for better memory later on (e.g., Glover, 1989; Whitten & Leonard, 1980). Such elaborative processing would be reduced for study trials relative to test trials, because the items are re-presented and thus immediately accessible.

The elaborative processing hypothesis can account for the lag, test type, and interference mediation effects by proposing that any manipulations designed to decrease the accessibility of an item during an intervening test will invoke more elaborative processing. The major weakness of the hypothesis, however, is the fact that it fails to specify the precise nature of this elaborative processing. In this way, it seems to provide more of a description rather than an explanation behind the pattern of results associated with the testing effect.

Variable Processing

The principle of encoding variability proposes that repeated items can be encoded in multiple ways, and if the number of different encodings for the same item is increased, so then are the number of potentially effective retrieval cues for that item (see Hintzman, 1976). Essentially, greater variability (i.e., different ways of encoding the same item) leads to greater retention. McDaniel and Masson (1985) proposed a similar hypothesis for the testing effect, in that intervening tests may increase the variable processing of tested items, thereby increasing the number of retrieval cues to those items and accounting for superior retention relative to non-tested items. McDaniel and Masson presented paired-associate items that were encoded based on semantic or phonemic association cues, followed by an intervening cued-recall test that utilized a semantic or phonemic cue, which was in turn followed by a final cued-recall test over all items that utilized a semantic or phonemic cue. Intervening tests were more beneficial to memory retention when a different type of cue was given during encoding and the final test (phonemic-semantic or semantic-phonemic), rather than when the same type of cue was given during encoding and the final test (phonemic-phonemic or semantic-semantic). Presumably, the intervening test increased the variability associated with the target item, making it more likely to be retrieved in the context of a different cue rather than the same cue at the time of the final test.

Cuddy and Jacoby (1982) tested participants' memory for target items after two intervening cued-recall tests in which the cues were either the same (e.g., *LAWYER: C _ _ RT* and *LAWYER: C _ _ RT*) or different (e.g., *LAWYER: C _ _ RT* and *LAWYER: _ OU _ _*) upon each encounter. On a final cued-recall test in which only the cue (e.g., *LAWYER:*

_____) was provided, more target items from the different cue condition were successfully recalled relative to items from the same cue condition.

The variable processing hypothesis is similar to, but more specific than, the elaborative processing hypothesis described above. It is possible that variable processing may represent a step toward defining the nature of the more vague *elaborative* processing. Manipulations designed to decrease the accessibility of an item at the time of an intervening test (such as increased lag, recall as opposed to recognition, and increased interference) also seem to decrease the probability that the item will be processed in the same way during the intervening test as it was during encoding. Instead, intervening test conditions that render items less accessible would increase the likelihood of *different* processing taking place during the intervening test relative to initial encoding, thereby increasing the number of potentially effective retrieval cues to tested items. Thus, in principle, the variable processing hypothesis can account for the lag, test type, and interference mediation effects.

Currently, however, the variable processing hypothesis lacks sufficient empirical data, and has not been fully developed to define the precise nature of the variability mechanism that presumably operates under test conditions more so than under study conditions. The next chapter reviews the concept of variability as it has been treated in the memory literature, and identifies the type of variable processing that is believed to operate in the testing effect. This unique type of variable processing, based on multiple cues sampled during retrieval, provides a specific operational definition of variable processing in the testing effect that allows for the formulation of testable hypotheses upon which the subsequent experiments are based. The three experiments designed to

investigate this variable processing perspective in the testing effect are outlined together in Chapter 3, and presented individually in the subsequent Chapters 4 (Experiment 1), 5 (Experiment 2), and 6 (Experiment 3). Finally, in the General Discussion (Chapter 7) the collective findings and implications of all three experiments are discussed.

CHAPTER II

VARIATIONS ON THE CONCEPT OF VARIABILITY

The basic premise of variable processing theory is that material learned in different ways increases the ways in which that material can be remembered later. Variable processing theories have been around for at least half a century, and have developed from general to more specific conceptualizations of the sources of variability in memory. Reviewing these theories is necessary in order to understand the principles behind exactly what is believed to induce variable processing in memory with respect to study conditions, as well as test conditions.

Stimulus Fluctuation Theory

The stimulus fluctuation theory (Estes, 1950, 1955a, 1955b) was perhaps one of the earliest well-developed theories proposing that variable learning conditions can promote retention. According to this theory (originally developed to account for stimulus-response learning), any given stimulus is made up of a specific, finite number of environmental features or *elements* (denoted as S). Upon encountering the stimulus, only a portion of the total elements in S are activated (denoted as s), whereas the remaining elements are not activated (denoted as s'). Upon repeated encounters with the same stimulus, an independent random sample of s is drawn from S . Thus, elements may fluctuate from a state of s to a state of s' , and vice versa, across repeated encounters with that stimulus.

The probability of making a response (denoted as R) is dependent upon the proportion of elements in s that are conditioned to (i.e., associated with) R . Long-term retention of R is directly related to the number of elements from s that have been associated with it, and the number of elements from s associated with R is maximized to the degree that each repeated encounter with a stimulus consists of a new independent sample from S . In other words, greater fluctuation of elements between s and s' during repeated stimulus encounters yields a higher probability of long-term retention for the response associated with that stimulus, because greater fluctuation increases the likelihood that different elements will become associated with R . The proportion of all the elements in S that are associated with R is positively related to the resistance of forgetting R over time. Estes' mathematical model of this theory has been successfully applied to spontaneous recovery effects (Estes, 1955b), as well as the distribution of practice (i.e., spacing) effect (Estes, 1955a).

Stimulus Sampling Theory

Stimulus sampling theory (Bower, 1972) is very similar to stimulus fluctuation theory, but emphasizes more of an active role of participants in the learning process. Instead of environmental elements tied to a specific stimulus, stimulus sampling theory proposes that variability takes place through *encoding operators*, which are processes performed on items during encoding that emphasize specific features (e.g., orthographic or semantic characteristics of verbal items). The same encoding operator can overlap with multiple items, to the degree that similar features emphasized by the encoding operator are shared by those items.

The probability of the specific encoding operators currently utilized depends not only on random fluctuation of context (i.e., the participant's moment-by-moment "stream of consciousness" (Bower, p. 93), but also on systematic fluctuations that call attention to specific features such as semantic (e.g., whether the word *court* is presented with the word *lawyer* or *tennis*) or orthographic (e.g., printing the letter *O* in red ink for the nonsense syllable *TOB*) characteristics. Like stimulus fluctuation theory, long-term retention associated with an item depends on the number of different encoding operators that have been associated with the item, such that a greater number of different encoding operators leads to greater retention of that item.

Stimulus sampling theory can account for not only improved retention on tests of recall or recognition, but also improved source memory such as list discrimination (i.e. remembering the specific lists in which items were presented), lag judgments, and frequency judgments (Bower, p. 95-115). The theory also points out that variability can be semantic or contextual in nature, which calls attention to two specific sources of variability that have been the focus of much of the empirical work on variable processing.

Semantic Variability

Variable processing explanations that focus on semantic variability propose that there can be more than one meaning, or interpretation, of the same stimulus, and long-term retention benefits to the degree that multiple meanings of that stimulus are processed during learning (Hintzman, 1976). Some evidence for this notion was obtained in a study by Soraci et al. (1999), who examined memory for homographs (ambiguous words with more than one meaning, e.g., *traffic jam* vs. *strawberry jam*) presented in the context of one or two meanings. Participants were given two definitions of the homographs that

either referred to the same meaning [BAT: (1) A winged mammal, (2) A nocturnal flyer] or to two different meanings [BAT: (1) A winged mammal, (2) A wooden club]. On a test of retention, more items were remembered that had previously been processed with two different meanings as opposed to one.

Contextual Variability

Variable processing explanations that focus on contextual variability propose that characteristics associated with an item's surrounding context are likely to change across repeated encounters with that item, and long-term retention benefits to the degree that multiple different contextual associations are acquired during learning (Anderson & Bower, 1972). One method of directly manipulating contextual factors in experimental paradigms is the use of context words in paired-associate learning. In this task, cue-target pairs (*speed: engine*) are presented multiple times, and upon each presentation, the cue (or context word) is either the same as before (*speed: engine*) or different (*valve: engine*). The to-be-learned material remains constant, but the contextual factors associated with it are either constant (when the same cues are used) or variable (when different cues are used). Young and Bellezza (1982; see also Madigan, 1969) presented cue-target pairs twice, with each presentation consisting of either same or different context words, and found greater free-recall retention for target items learned with different as opposed to similar context words.

Variable Processing during Retrieval

So far, all of the existing theories of variable processing have been formulated to account for the effects of variable presentations of items. However, it is not clear whether or not the same type(s) of variable processing are occurring during *retrieval* of items. If

variable processing accounts for the testing effect, then what is the nature of this variable processing, and why does it occur to a greater degree during retrieval (test trials) than during encoding (study trials)? There are at least two possible hypotheses.

1. Multiple Methods of Processing. One possibility is that tested items receive processing in the context of two different tasks (e.g., a study trial followed by a test trial), whereas studied items receive processing in the context of two similar tasks (e.g., a study trial followed by another study trial). Thus, the former may benefit memory more than the latter simply because two different processing methods have been used.

Although this hypothesis makes sense, there is empirical evidence to suggest that different as opposed to the same processing methods are actually less effective for retention. For example, Young and Bellezza (1982) had participants learn target items using one type of mnemonic device on two occasions (either the peg-word or story method), or using both types of mnemonic devices, one on each occasion. Recall of the items learned by two different mnemonic devices was actually impaired, relative to recall of the items learned by the same mnemonic device. Similarly, Green and Stillwell (1995) had participants learn target items by rating them on two different scales (variable condition), or rating them twice on the same scale (constant condition). Frequency discrimination of the items rated on two different scales was impaired relative to frequency discrimination of the items rated on the same scale twice.

Such a hypothesis also does not explain the test type mediation effect. If variable processing in the testing effect were the result of two different tasks utilized in test trials as opposed to study trials, then the type of intervening test should not matter, as long as it is different from the original study trial. The fact that intervening recall tests are more

beneficial than recognition tests (Carpenter & DeLosh, accepted pending revisions; Glover, 1989; Whitten & Bjork, 1977) suggests that the testing effect is not merely the result of two different processing methods.

2. *Multiple Cues.* A number of memory models propose that there are two processes involved in the recall of information from memory: first, an initial search phase in which potential target information is generated from memory, and second, a recognition/identification phase in which the potential target information is checked for accuracy (Anderson & Bower, 1972; Bahrick, 1970; Brainerd, Wright, & Reyna, 2002; Diller, Nobel & Shiffrin, 2001; Nobel & Shiffrin, 2001; Raaijmakers & Shiffrin, 1981).

The two processes involved in these models are described in great detail by the models of Shiffrin and colleagues, Search of Associative Memory (SAM, Raaijmakers & Shiffrin, 1981) and Assessment of Retrieval Completion-Retrieving Effectively from Memory (ARC-REM, Diller et al., 2001). During the first phase, memory is searched according to the information available in the cues that are provided.¹ Information based on the cue(s) that might be related to the target is sampled, then this information is subjected to the second process to determine whether or not that information is the correct target. If the information is not recognized as the correct target, then the search process continues and different information is sampled from the cue and again subjected to the recognition process. This process continues until either the correct target is sampled and recognized, or the correct target cannot be sampled and the decision is made to quit

¹ Memory cues are believed to always be present, even when no apparent experimenter-provided cues are provided. In tests of free-recall, for example, the cues consist of the participant's plan for how to search memory (e.g., "What was the first item I saw?" or, "What was an item from list 9?")

searching memory. A diagram that has been reconstructed from these two-process models is presented below in Figure 2-1. The ease with which the correct target is sampled depends on the strength of association between cue and target, such that greater cue-target strength facilitates retrieval and reduces the probability of returning to the first process for repeated sampling of information from memory.

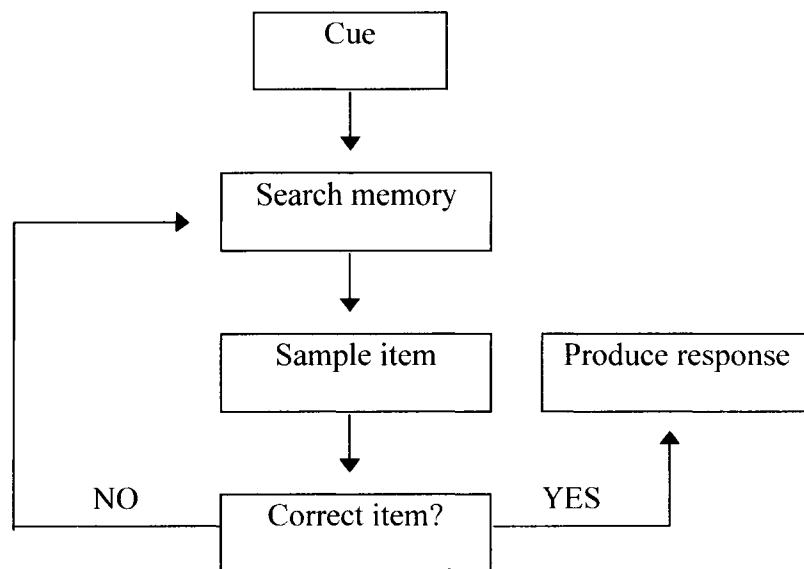


Figure 2-1. Reconstructed diagram of two-process model of retrieval.

If the search process ensues based on the information provided in the cue, and the probability of sampling the correct target is directly related to cue-target strength, then the correct target should be among the first information sampled in the case of a strongly related cue-target pair (e.g., *mother*: *child*), whereas the correct target should not be among the first information sampled in the case of a weakly related cue-target pair (e.g., *mitten*: *child*). All of the information activated that is associated with *mother* is more likely to contain the correct target word, *child*, relative to all of the information activated

that is associated with *mitten*. If the correct target is sampled and recovered in both cases, it is likely that extra information will have been sampled prior to the correct target when *mitten* as opposed to *mother* is used as the cue.

There is some evidence that the sampling of extra information prior to the correct target may benefit long-term retention of that target, by providing multiple cues for retrieval later on. For example, Soraci et al. (1994) examined the generation effect using a congruent cue (e.g., *An article of clothing: C _ P*) or an incongruent cue (e.g., *NOT a policeman: C _ P*), and found that the correct target item (*CAP*) was better remembered if it was generated in the context of an incongruent cue. Presumably, generating the correct target from an incongruent cue requires generation of the wrong response first (e.g., *COP*), which becomes attached to the memory of the correct response and acts as a future retrieval cue. Memory for the correct target generated from an incongruent cue includes more types of information (viz., information about what the item *is*, as well as what the item *is not*) than memory for the same target generated from a congruent cue (which only includes information about what the item *is*). Soraci et al. (1994; see also Soraci et al., 1999) proposed a *multiple cue hypothesis* to account for the incongruent generation effect, whereby multiple items generated during learning can serve as multiple retrieval cues later on.

According to the two-process models reviewed above, the search process that takes place during episodic memory retrieval (i.e., the testing effect) may be similar to the search process that takes place during semantic memory retrieval (i.e., the generation effect), in that items are sampled from memory in response to a cue, and some conditions (such as an incongruent or weakly related cue) increase the probability that extra

information will be sampled prior to the correct target. It is possible, therefore, that the multiple cue hypothesis might apply to the testing effect. During an intervening recall test, information based on the cue and related to the target is sampled from memory, and this information may provide retrieval cues to the target later on. During an intervening study opportunity, such information is not likely to be sampled and relied upon to the same degree, because the target item is currently presented. According to this reasoning, the relative advantage of tested over studied items should increase under intervening test trial conditions that promote the sampling of extra information prior to the correct target, and the relative advantage of tested over studied items should decrease under intervening study trial conditions that promote the sampling of extra information prior to the correct target. The next chapter outlines a series of experiments to determine whether and how this variable processing accounts for the testing effect.

CHAPTER III

OUTLINE OF EXPERIMENTS

Variable processing of semantic and/or contextual factors benefits retention for studied items, but it is unknown whether the same type of variable processing also benefits retention for tested items. This question was addressed in Experiment 1. Second, it is unknown whether variable processing in the form of multiple cues benefits retention for tested items but not studied items. This question was addressed in Experiment 2. Finally, it is unknown whether or not the amount of extra information sampled prior to the correct target is directly related to long-term retention, such that a larger quantity of extra information leads to greater retention. This question was addressed in Experiment 3.

Experiment 1

Experiment 1 was designed to answer the question of whether or not variable processing can benefit memory for both tested and studied items. Using a common manipulation of variable processing based on context words (e.g., Madigan, 1960; Young & Bellezza, 1982), a paired-associate design was used in which target words on both test and study trials were encountered with a context cue that was either the same or different from the one originally presented with the target. For example, in a study trial where the original cue-target pair was *toast: bread*, participants were presented with either the same target and cue again (*toast: bread*), or the same target but a different cue (*dough: bread*). On the test trials, participants were only presented with the cue and were required to

retrieve the target when the cue was either the same (*toast: _____*) or different (*dough: _____*) from the one encountered during original encoding.

If the testing effect generalizes to this procedure and these materials, then tested items should be better retained on a final free-recall test than studied items. Furthermore, if contextual variability manipulated by context words benefits memory for tested items as it does for studied items, then items encountered twice with different context words should be better retained than items encountered twice with the same context word, and this pattern should hold for both test and study trials. The specific effects of interest that are expected are: (1) a main effect for type of trial, whereby tested items are retained better than studied items, and (2) a main effect for type of cue, whereby different cues lead to better retention than same cues.

Experiment 2

Experiment 2 was designed to answer the more specific question of whether the type of variable processing proposed by the multiple-cue hypothesis (Soraci et al., 1994, 1999) applies to the testing effect. The same materials and basic design from Experiment 1 were used, except for the fact that the context words associated with the targets were always the same, and were either highly related to the target (*toast: bread*), or less related to the target (*basket: bread*). Target items are more accessible for tested items in the highly related condition because of the strong associations they share with the cues, whereas target items are less accessible for tested items in the unrelated condition because of the weaker associations they share with the cues (e.g., *basket* cues a number of other items besides *bread*). As such, the sampling of extra information prior to the correct target should be increased in the unrelated condition relative to the related condition.

If such extra information is sampled more often in the unrelated as opposed to related condition, and this extra information acts as retrieval cues later on, then tested items retrieved with unrelated cues should be better retained than tested items retrieved with related cues. Memory for studied items, on the other hand, would not be expected to vary as a function of the type of cue, because sampling extra information is less likely to take place and be relied upon in study trials to the same degree that it is in test trials. The specific effects of interest that are expected are: (1) a main effect for type of trial, whereby tested items are remembered better than studied items, and (2) a Trial x Cue Interaction, whereby retention for tested items is better for unrelated as opposed to related cues, but retention for studied items is no different whether they were presented with related or unrelated cues.

Experiment 3

Experiment 3 was designed to further test the multiple cue hypothesis by directly manipulating the quantity of extra information sampled prior to the correct target. This was accomplished by using the same cue-target pairs from Experiment 2 and a modification of a type of recognition task referred to as *wickelcall* (after Wickelgren & Corbett, 1977) by Nobel and Shiffrin (2001). In *wickelcall*, the cue member of a cue-target pair is presented and, after a delay, a potential target (either correct or incorrect) is presented along with the cue, and participants must indicate whether or not it is the correct target by a *yes* or a *no*. Nobel and Shiffrin (see also Diller et al., 2001) propose that the *wickelcall* task involves the same memory search and recognition processes as cued-recall, because memory search occurs during the first part of the procedure (when

only the cue is presented), and recognition occurs during the second part (when a potential target is presented).

Experiment 3 used a modified wickelcall task to simulate both the memory search and recognition processes that take place during retrieval, and to manipulate the amount of information sampled prior to the target during the memory search phase. Similar to wickelcall, the cue member of the cue-target pair was presented, however, between 0 and 3 potential targets (instead of just 1) were presented sequentially prior to the correct target. Participants were required to indicate whether or not the potential target was the correct one by a *yes* or a *no*. The correct target was always presented last, and either 0, 1, 2, or 3 incorrect targets preceded it. Once the correct target was presented and responded to, the test for that item was terminated and the next item was tested using the same procedure. Study trials also consisted of the presentation of the cue member of each cue-target pair, and the presentation of 0 through 3 incorrect targets prior to the correct target, but did not require participants to make recognition judgments about each target.

The modified wickelcall task was intended to simulate the retrieval process that occurs during cued-recall for correct targets that are accessed immediately (e.g., with 0 incorrect targets preceding it), versus correct targets that are accessed after extra information (e.g., 1, 2, or 3 incorrect targets) is sampled. If during tests of cued-recall, extra information is ordinarily sampled prior to the correct target and acts as retrieval cues in accordance with the multiple cue hypothesis, then target items that are presented and recognized after redundant extra information has been sampled (i.e., 1, 2, or 3 incorrect targets) should not be retained better than targets presented and recognized immediately (i.e., after 0 incorrect targets). In other words, if during tests of standard cued-recall,

participants are already sampling extra information of the type provided by the incorrect targets in the wickelcall task, then such extra information would be redundant and would not be expected to influence retention for tested items.

Ordinarily during study trials, such sampling of extra information is not as likely to occur, because the target is already presented. However, the current experiment is set up so that extra information *is* sampled on study trials. Because this information is not routinely sampled on study trials as it is on test trials, such information would not be redundant and so would be expected to influence retention for studied items. Specifically, providing extra cues in the form of incorrect targets for studied items should boost retention for studied items to the level of tested items when more incorrect targets are provided. Thus, the predicted effect of interest in Experiment 3 is a Trial x Number of Targets Interaction, such that retention for studied items improves with an increasing amount of incorrect targets, whereas retention for tested items is not affected by the number of incorrect targets. If multiple incorrect targets provide redundant cues for tested items but not for studied items, then the relative advantage for tested over studied items is expected to occur most strongly at 0 incorrect targets, and this relative advantage should decrease as the number of incorrect targets increases.

Key Design Features

The three experiments in this study were carefully designed to address the research questions described above, while at the same time control for some methodological concerns that have been apparent in past research on the testing effect. Specifically, four elements of the experimental design are worth noting. First, the study trials in all three experiments consisted of a task that required the active processing of

relational information between cue and target. In some past studies of the testing effect, the study trials consisted of a re-exposure to the items without any apparent processing activity applied to these items (e.g., Cull, 2000; Nungester & Duchastel, 1982). One potential concern, therefore, is that studied items are given less attention, or even ignored, relative to tested items simply because no processing activity is required of studied items that are simply re-exposed. Other studies have attempted to equate the activity going on during test and study trials by requiring participants to read the studied items aloud and recall the tested items aloud (Allen et al., 1969; Carrier & Pashler, 1992; Kuo & Hirshman, 1996, 1997). Although this effectively matches the activity associated with the response for test and study trials, one might still argue that tested items are generally processed to a fuller extent because of the activity associated with retrieval, and if studied items were given some additional processing, they would be retained as well as tested items. The current study addressed this concern by utilizing a study trial that required relational processing between cue and target, thereby insuring that participants would perform some processing activity on the studied items. If the testing effect is still obtained under these conditions, then it is possible to rule out an artifactual explanation based on a heightened degree of general processing activity (i.e., attention) for tested items at the expense of studied items.²

² Previous studies from our lab have also used a study trial that requires entry of the same response as on test trials, only in the study trials the response is entered by copying it, whereas in the test trials the response is entered by retrieving it. A significant testing effect emerges under these conditions, ruling out an explanation based on the activity associated with entering a response (see Carpenter & DeLosh, accepted pending revisions).

Secondly, all three experiments manipulated the type of trial (test versus study) in a between-list fashion, in which all of the items on a list were either presented as a test or study trial. Some concern has been raised for studies that manipulate the type of trial within-lists (e.g., in which some of the items from the list are presented as a test trial, and other items from the same list as a study trial). In a within-list design, it is possible that participants could be devoting rehearsal time to tested items at the expense of studied items, based on a subjective judgment that studied items are more accessible and well-learned relative to tested items. This *displaced rehearsal hypothesis* has been proposed to account for memory phenomena that are similar to the testing effect, such as the spacing effect (Green, 1989, 1990) and the generation effect (Slamecka & Katsaiti, 1987), so there is reason to be concerned that such a hypothesis might also play a role in the testing effect. There is evidence against the displaced rehearsal hypothesis to explain the testing effect, based on the emergence of a significant testing effect using both within- and between-list designs (Carrier & Pashler, 1992). Nonetheless, the current study reduced the possibility of displaced rehearsal by presenting all test and study trials in a between-list design. If the testing effect still exists in a between-list design, this would add support to the reliability of the testing effect regardless of list composition, and thus rule out an explanation based on displaced rehearsal.

Finally, the same target items, type of intervening test, and type of final test were used for all three experiments. This made it possible to vary the experimental manipulations and observe key changes in retention for tested versus studied items without the potential extraneous influences of different experimental methods and materials. Essentially, this allows the testing effect to be investigated with the control and

precision necessary to attribute the results to the experimental manipulations and not to any artifacts associated with specific items or type of tests.

CHAPTER IV

EXPERIMENT 1

The purpose of Experiment 1 was to determine whether variations in context benefit memory for both tested and studied items. A paired-associate design was used in which context words were presented to the left of to-be-remembered target words (e.g., *toast: bread*). All of the target words were encountered twice, once at initial encoding, and once again for either another study trial (ST) in which cue and target were re-presented, or for a cued-recall test trial (TT) in which only the cue was presented and the target had to be retrieved (e.g., *toast: _____*). Contextual variability was manipulated by varying the context word associated with the target word between the first and second encounters. In order to accomplish this, both TT and ST consisted of either the same context word that was originally encoded (e.g., if *toast: bread* was encoded, then *toast: _____* was displayed for TT, and *toast: bread* for ST), or a different context word from that which was originally encoded (e.g., if *toast: bread* was encoded, then *dough: _____* was displayed for TT, and *dough: bread* for ST). Thus, the target word remained the same, but the context word that was paired with each target word on TT and ST was either the same or different from the one encountered during the original encoding episode. The effects of type of trial (TT vs. ST) and type of context word (same vs. different) were evaluated using a final free-recall test over all of the target words.

It was hypothesized that the contextual variability associated with target items would be greater when different as opposed to the same context words were used across

the first and second encounters with each target item. If contextual variability benefits retention for both TT and ST, then more items should be recalled on the final test that were presented with different as opposed to the same context words. Specifically, it was expected that retention would be superior for TT as opposed to ST, and for different cue as opposed to same cue conditions.

Method

Participants

Twenty-two male and 45 female undergraduate students volunteered to participate in order to fulfill partial requirements for an introductory psychology course at Colorado State University. Participants were randomly assigned to one of the four experimental conditions (described in the *Design* section) and were tested individually on personal computers.

Materials

Word lists were constructed using a pool of 64 nouns that were selected based on the characteristics of word frequency, concreteness, number of syllables, and word length. Wilson's (1988) database was used to select items with a frequency of occurrence of greater than 20 per million, and a concreteness rating from 500 to 700. Each item consisted of one to two syllables, and ranged between five and seven letters. Eight lists were constructed, each consisting of eight items that were orthographically, phonologically, and semantically distinct.

Once all of the target items were selected, the norms of Nelson, McEvoy, and Schreiber (1998) were used to obtain two different cues, each with comparable cue-to-target strength, for each target. Cue-to-target strength (i.e., the probability of producing

each target given the cue) ranged from .20 to .50. This range of probabilities was believed to be strong enough to facilitate production of the correct target from memory when given either cue, while at the same time weak enough to reduce the probability of correctly guessing the target. Word lists were constructed so that neither of the two cues associated with each target could also be associated with any other targets in the same list. For example, the two target items *train* and *music* could share the common cue *conductor*, so items such as these were always placed in different lists. A complete list of the target items, cues, and cue-to-target strength, is reported in Appendix 1.

Design

During the initial encoding phase of the experiment, each target item was presented along with one of its cues, and participants were required to rate the degree of relatedness between the two, using a five-point scale in which larger numbers indicated a greater degree of relatedness. Participants were instructed to remember only the target item, and to try and use the rating between cue and target to help them remember it. A sample computer screen for the encoding task is shown below in Figure 4-1.

Rate the relatedness between the two items below.
Use this to help you remember the underlined word.

toast bread

1.....2.....3.....4.....5

Not related Somewhat related Highly related

Figure 4-1: Sample computer screen for encoding task in Experiments 1-3.

All eight target items, along with one of their cues, were presented and rated in this fashion. Two of the eight lists were designated as practice lists that were administered at the beginning of the experiment, and these were not included in the analysis. Of the six remaining lists, three were designated as ST and three as TT. For the ST, the same encoding task was again presented and participants were required to rate the relatedness between cue and target to help them remember the target word. For the TT, only the cue was presented and participants were required to type in the correct target word.

For half of the TT and ST, the cue displayed was the same one that was utilized during the encoding task, whereas for the other half, it was different from the one utilized during the encoding task. Thus, the following four conditions (illustrated on the next page in Figure 4-2) were used in Experiment 1: (1) TT-same cue, (2) TT-different cue, (3) ST-same cue, and (4) ST-different cue. Four different versions of the experiment were created (see Appendix 2) to ensure that all items appeared equally often in each of the four conditions. The designation of each list to each condition was random.

Participants were given all eight lists in the same order, and the presentation of items within each list was random for each participant. The measure of retention was a final free-recall test administered at the end of the experiment, which required participants to type in all of the words they could remember from the six experimental lists. A free-recall test represents the most accurate assessment of variable processing during learning, because a variety of different contextual cues must be relied upon (Green, 1989). Participants in all four versions of the experiment completed the same final test.

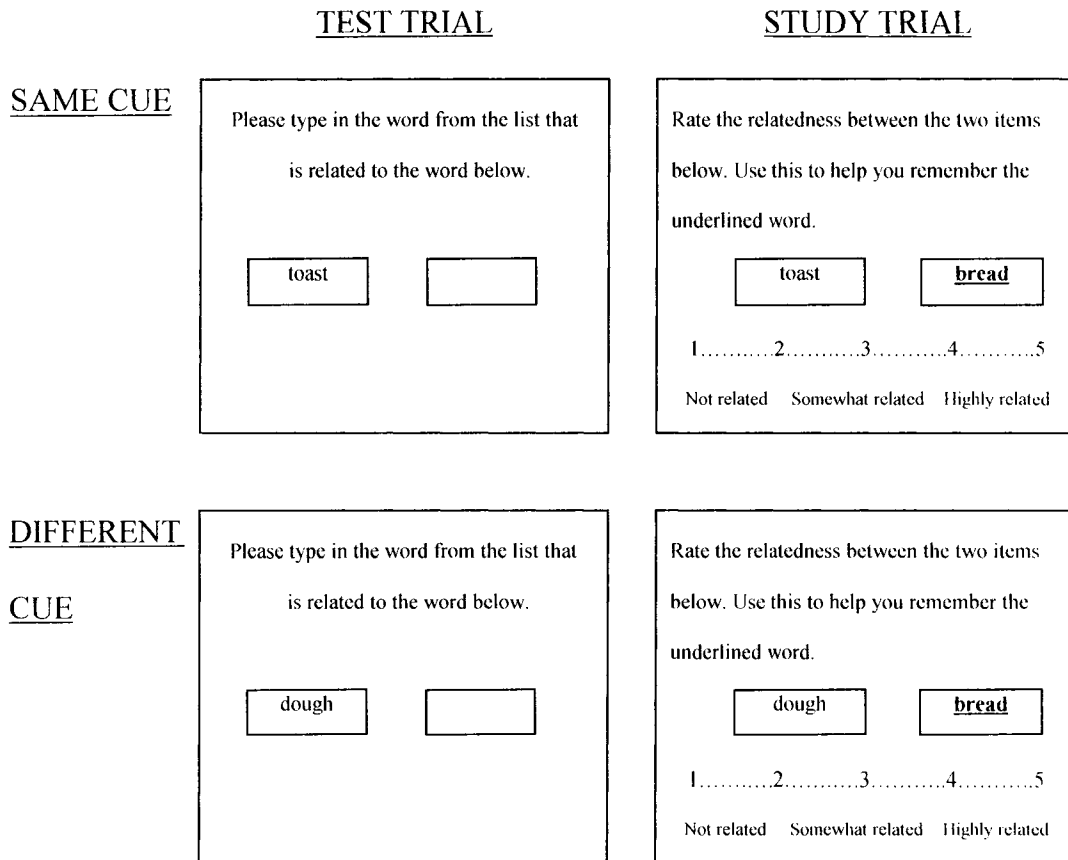


Figure 4-2. Sample computer screens for TT and ST with same and different cues. Original encoding cue was *toast*.

Procedure

Participants first read instructions on the computer monitor, then were given two practice lists (one TT and one ST) to demonstrate the procedure. The instructions informed participants that they were required to remember only the underlined (i.e., target) word, and the cue that was presented with the underlined word may or may *not* be the same in the TT and ST as it was in the encoding task.

During the encoding task, each cue-target pair in a list was presented for an indefinite amount of time, and participants indicated the degree of relatedness between the two words by pressing the numerical keys 1 through 5 on the keyboard. Upon each

key press, a new cue-target pair appeared. After all eight cue-target pairs were presented and rated in this fashion, eight single-digit numbers were presented sequentially at a rate of 1000 ms per number and 1000 ms inter-stimulus interval, for a total of about 15 s. The numbers constituted a distracter task in which participants were required to add the numbers together and record the correct total by typing it into the computer. After all eight numbers had been shown, participants were prompted to type in the total by the instructions “Please type in the sum total of the numbers you just saw, followed by the ENTER key.”

Following the distracter task, participants were given either a TT or ST over the list that was just presented. For the TT, participants were presented with the cue (either same or different from the one encountered during the encoding task) and were required to type in the correct target word, followed by the ENTER key, or just the ENTER key alone if they were unable to retrieve the correct target word. The cue was presented in the left-hand box, and the target word that participants typed in appeared in bold, underlined font in the right-hand box (Figure 4-2). No feedback was given to verify whether participants’ responses were correct.

For the ST, participants were presented with both target and cue (either same or different from the one encountered during the encoding task), and were required to rate the relatedness between the two words using the same five-point scale from the encoding task. As in the encoding task, rating the relatedness between the two words in the ST was participant-paced. Following completion of the TT or ST, participants were presented with the next list, and the same procedure was repeated.

Following the presentation of all eight word lists, participants were given a five-minute distracter task in which they were required to type in the names of as many U. S. states that they could think of. The computer screen was numbered 1 through 50, and each state participants typed in appeared next to a number so that they would be aware of how many states they remembered. After five minutes, the states and numbers were cleared from the computer screen, and instructions for the final test were displayed. These instructions informed participants that they would be required to type in as many of the target words from the experiment that they could remember, and to press the spacebar when they were ready to begin.

For the final test a blank computer screen was displayed, and participants were instructed to type each target word one at a time, followed by the ENTER key. Each word that participants typed appeared in the center of the screen, and disappeared upon pressing the ENTER key, so that each new word was typed onto a blank screen. In this way, participants retrieved each item without reviewing previously-retrieved items. A maximum of ten minutes was allotted to complete the final test, after which participants were instructed by the experimenter to stop. After completion of the experiment, participants were fully debriefed as to the purpose of the study, how the data would be used, and the expectations and implications for the results. The entire procedure lasted approximately 35 minutes.

Results

Intervening Test Accuracy. The proportion of items correctly recalled on the intervening tests with same and different cues was computed. Using a paired-samples t-test, a significant difference was observed [$t(66) = 12.43, p < .05$] between the mean

accuracy rate for TT with the same cues ($M = .95$, $SD = .09$) versus TT with different cues ($M = .66$, $SD = .17$), verifying, as expected, that it was harder to retrieve the correct target word when the test consisted of a cue different from the one encountered during encoding.

Final Retention for TT versus ST with Same versus Different Cues. The proportion of items correctly recalled on the final test from all four conditions (see Figure 4-2) was computed. A 2 x 2 (Trial x Cue) repeated measures Analysis of Variance (ANOVA) indicated a significant main effect for trial [$F(1, 66) = 73.42$, $p < .05$, $MSE = .009$, partial $\eta^2 = .53$], a non-significant main effect for cue [$F(1, 66) = .004$, $p > .05$, $MSE = .015$], and a significant Trial x Cue Interaction [$F(1, 66) = 6.10$, $p < .05$, $MSE = .013$, partial $\eta^2 = .09$]. The main effect of trial indicates that tested items were retained significantly better than studied items, and the Trial x Cue Interaction indicates that the rate of retention for tested versus studied items differed depending on whether the cue was same or different. Based on the means and standard deviations for all conditions (displayed on the next page in Table 4-1), it appears that the interaction effect was driven by an advantage for same versus different cues on TT, and a disadvantage for same versus different cues on ST (see Figure 4-3 on the next page).

What appears to be a lack of effect for variable processing on tested items is most likely an artifact of the low retrieval rates for tested items in the different-cue condition. Tested items in the same versus different cue conditions were retrieved at rates of .95 and .66, respectively. It is well-known that low retrieval rates on intervening tests can lead to underestimates of experimental manipulations, because items that are not successfully retrieved on one intervening test without feedback are highly unlikely to be retrieved on a

final test (Bjork, 1988; Glover, Krug, Hannon, & Shine, 1990; Kuo & Hirshman, 1996, 1997; Runquist, 1983, 1986).

Table 4-1.

Type of Cue	Type of Trial		Total
	Test	Study	
Same	.30 (.15)	.17 (.14)	.23 (.15)
Different	.27 (.15)	.20 (.15)	.23 (.15)
Total	.29 (.15)	.18 (.14)	

Note. Standard deviations are given in parentheses.

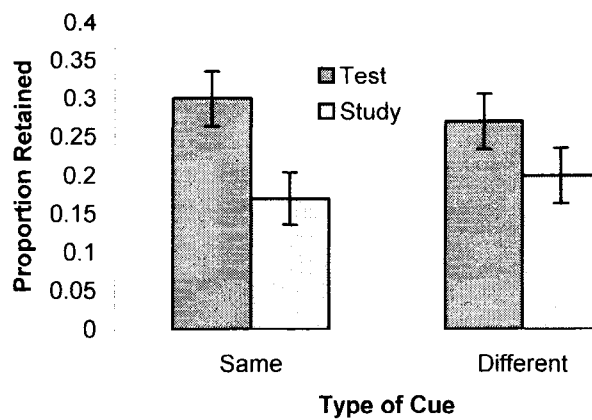


Figure 4-3. Proportion of items retained on final test as a function of type of trial and type of cue. Note: Error bars represent 95% confidence intervals.

There are a number of ways of examining data with the influence of such an artifact removed. One way is to limit the data to only those participants who successfully retrieved tested items in both the same and different cue conditions at a sufficiently high and comparable rate so as not to bias final test performance in favor of the same cue condition at the expense of the different cue condition. The data were limited to those participants ($n = 33$) whose retrieval rates were greater than .66 (the mean rate for the different cue condition) for both same and different cue test conditions. Examination of these data reveals trends that are more consistent with the variable processing effect, whereby different cues are consistently superior to same cues (see top half of Table 4-2 in Appendix 3). When data are limited to these 33 participants, there is both a testing effect and a variable processing effect. The testing effect remains, and the variable processing effect increases, when the data are further limited to those participants ($n = 14$) whose intervening retrieval rates are even more compatible (greater than .75, see bottom half of Table 4-2 in Appendix 3). The descriptive and inferential statistics for these analyses are reported in Appendix 3.

Another way to remove the artifact of low intervening test retrieval rates is to conditionalize final test performance upon successful performance during the intervening tests. Conditionalizing scores in this way helps to remove the final test performance bias in favor of study trials due to low intervening test performance, while at the same time does not require the elimination of any data. The standard unconditionalized data reported above (Table 4-1) were derived by calculating the number of items recalled on the final test from each of the four conditions, divided by the total number of items in each condition, regardless of whether or not these items were correctly recalled on the

intervening test. Conditionalized data, on the other hand, are derived by calculating the number of items recalled on the final test from each of the four conditions, divided by the total number of items in each condition that were successfully entered on the intervening test. For example, if there are 16 items assigned to the test trial condition with different cues, and the participant successfully recalls 10 of these on the intervening test and 5 on the final test, then the proportion retained according to the unconditionalized score would be $5 / 16 = .31$, whereas the proportion retained according to the conditionalized score would be $5 / 10 = .50$.

Final retention scores for all four conditions were conditionalized in this way, and the means and standard deviations are reported on the next page in Table 4-3 (see also Figure 4-4 on the next page). The 2 x 2 (Trial x Cue) Repeated Measures ANOVA was re-run on the conditionalized data, and yielded a significant main effect for type of trial [$F(1, 66) = 80.71, p < .05, MSE = .035, \text{partial } \eta^2 = .55$], a significant main effect for type of cue [$F(1, 66) = 12.22, p < .05, MSE = .035, \text{partial } \eta^2 = .16$], and a significant Trial x Cue Interaction [$F(1, 66) = 4.86, p < .05, MSE = .027, \text{partial } \eta^2 = .07$]. Final test performance was superior for tested items relative to studied items, and for items presented with different cues relative to items presented with the same cues. The interaction effect, which was significant but rather small in strength (with an effect size of only .07), appears to be driven by a stronger advantage for different versus same cues with tested items, compared to a relatively weaker advantage for different versus same cues with studied items. Nonetheless, different cues benefited memory more than same cues for both tested and studied items.

Table 4-3.

Type of Cue	Type of Trial		Total
	Test	Study	
Same	.32 (.16)	.17 (.14)	.24 (.16)
Different	.44 (.31)	.20 (.15)	.33 (.26)
Total	.37 (.24)	.18 (.14)	

Note. All scores are based on conditionalized data. Standard deviations are given in parentheses.

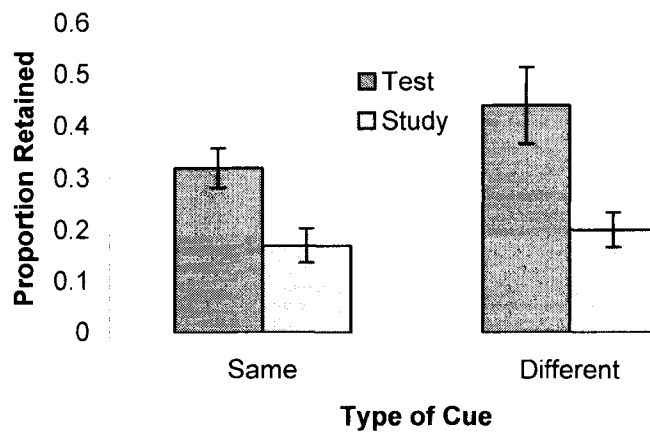


Figure 4-4. Proportion of items retained on final test as a function of type of trial and type of cue. Note. All scores are based on conditionalized data. Error bars represent 95% confidence intervals.

This re-examination of the data reveals consistent testing effects, as well as significant variable processing effects for both tested and studied items, when the intervening test retrieval rates for variable (different cue) and constant (same cue) conditions are comparable.

Discussion

A significant testing effect emerged under conditions of both same and different cues, using either conditionalized and unconditionalized scores. This is consistent with a number of prior studies showing that memory for tested items is superior to memory for studied items even when no corrective feedback is provided (Allen et al., 1969; Cull, 2000; Darley & Murdock, 1971; Nungester & Duchastel, 1982; Petros & Hoving, 1980; Wheeler & Roediger, 1992). The current experiment provides evidence that the testing effect generalizes to paired-associate learning when the cues given on the intervening test are either the same or different from those encountered during encoding.

A significant effect for variable processing also emerged for conditionalized scores, such that tested and studied items presented with a different cue were retained better than those presented with the same cue. These results are consistent with past studies investigating variable processing using same versus different cues for studied items (Madigan, 1969; Young & Bellezza, 1982), and also extend on this past research in demonstrating that the same benefit applies to tested items, provided that the intervening retrieval rates are not too low. The results for tested items are consistent with those of Cuddy and Jacoby (1982), who investigated the effects of similar versus different cues for tested items and also found that a different cue presented between encoding and the

intervening test benefited memory more so than when the same cue was presented between encoding and the intervening test.

The results of Experiment 1 answer the fundamental question of whether or not tested items benefit from a well-known manipulation of variable processing based on same versus different cues. The next experiment was conducted in order to determine whether extra information sampled during retrieval might be a source of variability, unique to tested as opposed to studied items, that could explain the advantage in final retention for the former over the latter.

CHAPTER V

EXPERIMENT II

The purpose of Experiment 2 was to determine whether the multiple-cue hypothesis (Soraci et al., 1994, 1999) might apply to the testing effect. Similar to the studies of Soraci and colleagues on the generation effect, the current study used cue-target pairs and manipulated the degree of relatedness between cue and target, such that the probability of producing the target given the cue was either high (e.g., *toast: bread*) or low (e.g., *basket: bread*). In the first condition, the correct target is likely to be produced more quickly and the probability of sampling extra information prior to the correct target is reduced relative to the second condition. However, memory of a target item retrieved from a highly related cue, relative to an unrelated cue, may suffer at the time of a final test because direct access to the item during the intervening test did not necessitate the sampling of extra information that could provide later retrieval cues.

Method

Participants

Thirty-four male and 21 female undergraduate students volunteered to participate in order to fulfill partial course requirements for an introductory psychology course at Colorado State University. Participants were randomly assigned to one of the four conditions (described in the *Design* section) and were tested individually on personal computers.

Materials

The target items used in Experiment 2 were the same as those used in Experiment 1. In addition, one set of cues was taken from Experiment 1 to serve as the related cues (with cue-to-target strength ranging from .20 to .50), and a different set of cues was sampled from the norms of Nelson et al. (1998) to serve as the unrelated cues (with cue-to-target strength of .01 in all cases). Thus, for the related cues, 20 to 50% of the participants in Nelson et al.'s norming study produced the target when given the cue, whereas for the unrelated cues, only 1% of the participants produced the target when given the cue. All of the target items, cues, and cue-to-target strength are presented in Appendix 1.

Design and Procedure

The design and procedure were very similar to Experiment 1, except that all of the cues presented on test trials (TT) and study trials (ST) were the same as those provided during the encoding task, and the cues were either related or unrelated to the targets. Thus, the following four conditions (see Figure 5-1) were used in Experiment 2: (1) TT-related, (2) TT-unrelated, (3) ST-related, and (4) ST-unrelated. Four different versions of the experiment were created to ensure that all items appeared equally often in each of the four conditions (see Appendix 2). In all other respects, Experiment 2 was identical to Experiment 1. The entire procedure lasted approximately 35 minutes.

Results

Intervening Test Accuracy. The proportion of items correctly recalled on the intervening tests with related and unrelated cues was computed. Using a paired-samples t-test, a significant difference was observed [$t(54) = 3.60, p < .05$] between the mean

accuracy rate for TT with related cues ($M = .97, SD = .07$) versus TT with unrelated cues ($M = .90, SD = .17$), verifying, as expected, that it was easier to retrieve the correct target word when the test consisted of a cue that was more associated with the target. Both intervening retrieval rates were sufficiently high in Experiment 2, to reduce the concern for artifactually deflated estimates of the effect of cue manipulation on tested items.

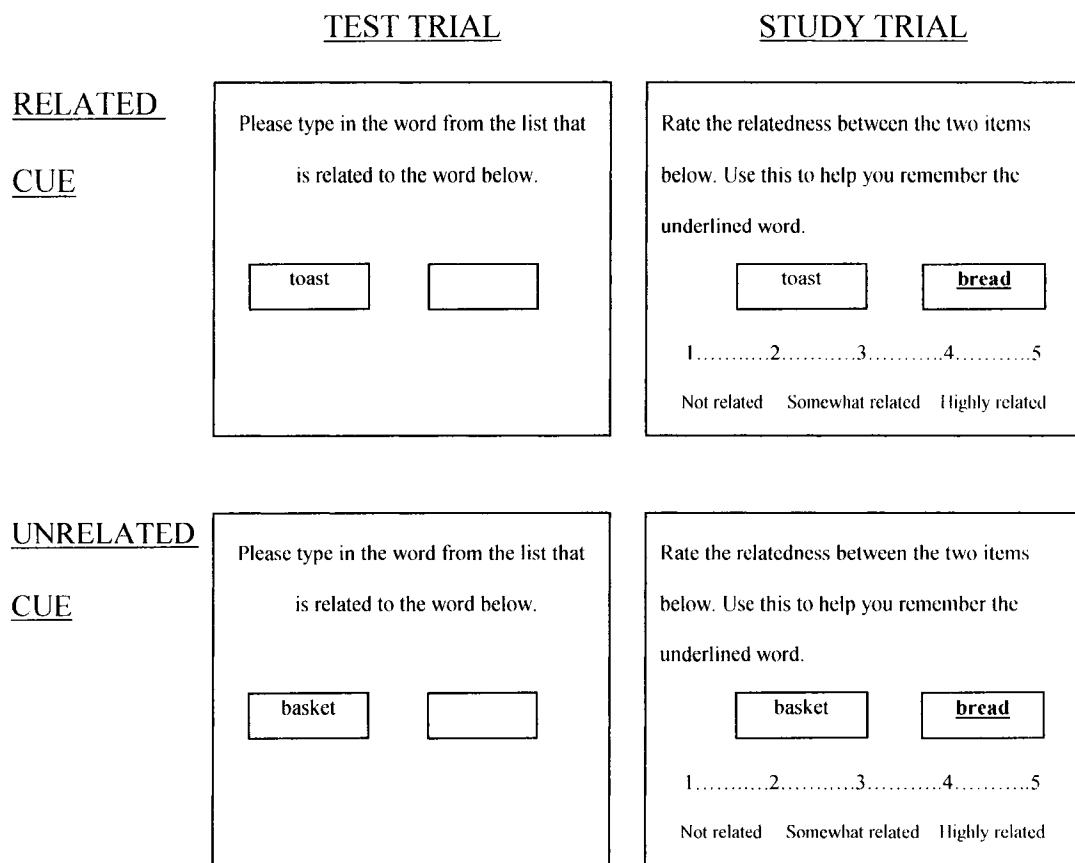


Figure 5-1. Sample computer screens for TT and ST with related and unrelated cues.

Final Retention for TT versus ST with Related versus Unrelated Cues. The proportion of items correctly recalled on the final test from all four conditions (see Figure 5-1) was computed. A 2 x 2 (Trial x Cue) repeated measures Analysis of Variance

(ANOVA) indicated a significant main effect for trial [$F(1, 54) = 133.09, p < .05, MSE = .01, \text{partial } \eta^2 = .71$], a non-significant main effect for cue [$F(1, 54) = 2.02, p > .05, MSE = .02$], and a significant Trial x Cue Interaction [$F(1, 54) = 4.23, p < .05, MSE = .02, \text{partial } \eta^2 = .07$]. The main effect of trial indicates that tested items were retained significantly better than studied items, and the Trial x Cue Interaction indicates that the rate of retention for tested versus studied items differed depending on whether the cue was related or unrelated. Based on the means and standard deviations for all conditions (displayed below in Table 5-1), it appears that the interaction effect was driven by an advantage for unrelated as opposed to related cues on TT, and no advantage for either type of cue on ST (see also Figure 5-2).

Table 5-1.

<u>Type of Cue</u>	Type of Trial		Total
	Test	Study	
Related	.23 (.16)	.11 (.12)	.17 (.15)
Unrelated	.30 (.19)	.10 (.10)	.20 (.18)
Total	.27 (.18)	.11 (.11)	

Note. Standard deviations are given in parentheses.

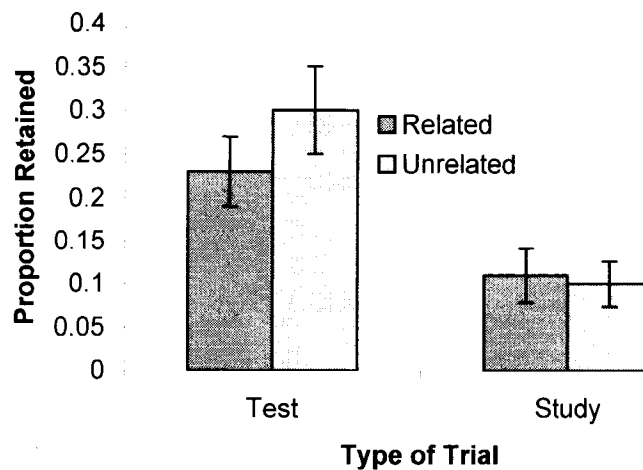


Figure 5-2. Proportion of items retained on final test as a function of type of trial and type of cue. Note: Error bars represent 95% confidence intervals.

Discussion

A significant testing effect emerged under conditions of both related and unrelated cues, consistent with a dense body of literature on testing effects (Allen et al., 1969; Cull, 2000; Darley & Murdock, 1971; Nungester & Duchastel, 1982; Petros & Hoving, 1980; Wheeler & Roediger, 1992). The current study provides evidence that the testing effect generalizes to paired-associate learning when the cues given on the intervening test are either related or unrelated to the targets.

The significant Trial x Cue Interaction indicates that unrelated cues were more beneficial than related cues for test trials, but not for study trials. These results support the prediction that tested items would be better retained if tested by unrelated as opposed to related cues, presumably because unrelated cues promote sampling of extra information

prior to the correct target, and this extra information may act as retrieval cues later on. Studied items would not be expected to benefit from unrelated cues, because the sampling of information from memory to act as future retrieval cues is not likely to occur when the target item is readily available.

The results of Experiment 2 support an explanation for the testing effect based on the multiple cue hypothesis. However, the experiment did not directly test whether and how much extra information was sampled for test trials with unrelated as opposed to related cues. It was assumed, but not verified, that the unrelated cues would promote sampling of extra information that would act as retrieval cues, but the possibility of an alternative explanation cannot be effectively ruled out. For example, it is possible that unrelated cue-target pairs engender more distinctive or elaborate processing relative to related cue-target pairs, because the connection between the items is more remote. Experiment 3 provided a more direct test of the multiple cue hypothesis in the testing effect, by using the same items from Experiment 2 and controlling for the type and amount of information sampled prior to the correct target.

CHAPTER VI

EXPERIMENT III

The purpose of Experiment 3 was to determine whether the multiple cue hypothesis (Soraci et al., 1994, 1999) accounts for the testing effect when the specific type and amount of extra information sampled prior to the correct target is controlled. Using the unrelated cue-target pairs from Experiment 2, a modified version of the wickelcall task (Nobel & Shiffrin, 2001; Wickelgren & Corbett, 1977) was used in which the cue member of each cue-target pair (e.g., *basket: _____*) was presented briefly, and the correct target was either presented first or after 1, 2, or 3 incorrect targets. This aspect of the design is illustrated below in Table 6-1, where the correct target is underlined.

Table 6-1.

Extra Targets	Items presented				
	Cue	Cue: Target 1	Cue: Target 2	Cue: Target 3	Cue: Target 4
0	basket: _____	basket: <u>bread</u>			
1	basket: _____	basket: wicker	basket: <u>bread</u>		
2	basket: _____	basket: wicker	basket: fruit	basket: <u>bread</u>	
3	basket: _____	basket: wicker	basket: fruit	basket: case	basket: <u>bread</u>

Upon each cue-target presentation, participants were required to indicate whether or not the word on the right-hand side was the correct target. The task was designed to simulate the sampling and recognition processes that occur during retrieval, by providing sampled information followed by a recognition decision on that information. This sampled information was operationally defined as the number of potential targets activated prior to the correct target. If tested items are ordinarily retained better than studied items because extra information of a similar type to these incorrect targets is sampled during retrieval, then providing such extra information for tested items would serve as redundant processing which would not be expected to benefit retention. If studied items are less well-retained than tested items because such extra information is not ordinarily processed on study trials, then providing extra information for studied items would *not* serve as redundant processing and thus would be expected to benefit retrieval. Specifically, it is expected that retention for tested items would be comparable at all levels of incorrect targets, whereas retention for studied items would increase when more incorrect targets are provided.

Method

Participants

Thirty-eight male and 40 female undergraduate students volunteered to participate in order to fulfill partial course requirements for an introductory psychology course at Colorado State University. Participants were randomly assigned to one of the eight conditions (described in the *Design* section) and were tested individually on personal computers.

Materials

In addition to the unrelated cue-target pairs from Experiment 2, three new targets for each cue were sampled from the norms of Nelson et al. (1998), to yield a total of four targets for each cue. All of the new targets sampled consisted of cue-to-target strength that ranged from .01 to .05 (comparable to the already-existing targets at .01). Two exceptions to this occurred: the cue items *pupil* and *scissors* had limited numbers of targets, so cue-to-target strength was more variable for the extra targets associated with these items. For *pupil*, the cue-to-target strength for the three new targets ranged from .08 to .68. For *scissors*, only two targets were available (cue-to-target strength ranging from .02 to .87), so a third target (*sew*) was chosen by the experimenter. The cues and all four targets, along with cue-to-target strength, are illustrated in Appendix 4.

Design

The experiment represents a 2 x 4 (Type of Trial x Number of Extra Targets) factorial design, in which type of trial (test versus study) and number of extra targets (0, 1, 2, or 3) were both manipulated within-subjects. Half of the six lists were presented as test trials (TT), and half as study trials (ST), and two of the items in each list were presented with 0 through 3 incorrect targets. For TT, the cue alone was presented, followed by the target after 0, 1, 2, or 3 incorrect targets, and participants had to indicate whether or not each target presented was the correct target by pressing the Y or N key to indicate yes or no, respectively. For ST, the cue was also presented, followed by the target after 0, 1, 2, or 3 incorrect targets, and participants had to rate the relatedness of each cue-target pair using the same five-point scale from the encoding task.

Eight different versions of the experiment were created so that all items appeared equally often as TT or ST, and with 0 through 3 incorrect targets preceding the correct target. Number of incorrect targets was manipulated within-lists, with a random sample of two items from each list presented with 0, 1, 2, or 3 incorrect targets. Each item was presented randomly during TT and ST, so there was no predictable order to the number of incorrect targets associated with each correct target. A full illustration of all eight experimental versions is displayed in Appendix 5. The same free-recall final test from Experiments 1 and 2 was used to assess retention at the end of the experiment.

Procedure

Participants first read instructions on the computer monitor, then completed two practice lists (one TT and one ST, each with 0 through 3 incorrect targets) to demonstrate the procedure. The instructions informed participants that they would be presented with some incorrect target words during TT and ST, and to try and distinguish the correct target from the incorrect ones.

The encoding task took place in the same fashion as in Experiments 1 and 2, and the same number-adding distracter task took place directly after the encoding task. After the distracter task, a TT or ST was presented over the most recently-presented list. For the TT, participants were presented with the cue alone for 1000 ms, followed by each of the targets, presented one at a time, for an indefinite amount of time. Upon each target presentation, participants were instructed to press the Y or N key to indicate yes or no, respectively, as to whether or not the presented target was the correct one. Upon each Y or N key press, either the next item from the list was tested by presentation of a new cue (in the 0 incorrect target condition), or the next potential target for the same cue appeared

(for items in the 1, 2, or 3 incorrect target conditions). After presentation of the correct target and the participant's response, the next item from the list was tested by the presentation of a new cue. After all of the list items had been tested in this fashion, the next list was presented for encoding using the five-point rating scale.

For ST, participants were presented with the cue alone for 1000 ms, followed by each of the targets, presented one at a time, for an indefinite amount of time. Upon each target presentation, participants were instructed to rate the relatedness of each cue-target pair by pressing the key 1 through 5 according to the same five-point scale from the encoding task. Upon each key press, either the next item from the list was presented with a new cue (in the 0 incorrect target condition), or the next potential target for the same cue was presented to be rated (for items in the 1, 2, or 3 incorrect target conditions). The correct target was always presented in bold, underlined font so that participants would be aware of which one it was. After presentation of the correct target and the participant's response, the next item from the list was presented with a new cue. After all of the list items had been presented in this fashion, the next list was presented for encoding using the five-point rating scale. Figure 6-1 on the next page illustrates sample computer screens for all of the conditions in Experiment 3.

Following the last TT or ST, the same final test distracter task from Experiments 1 and 2 was used, in which participants were required to type in the names of as many U. S. states that they could think of in five minutes. Following the distracter task, instructions for the final free-recall task were displayed, informing participants that they must type in the names of as many target items that they could remember from the entire experiment. As in Experiments 1 and 2, the final test screen was blank, and responses were typed in

		Number of Incorrect Targets											
Trial Type		Cue	0		1		2			3			
Study													
		Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	Please rate the relatedness between the two items below. 1.....2.....3.....4.....5	
Test													
		Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	Please indicate whether the item below is the correct target. Press Y for yes and N for no.	
Correct Answer on Test Trials			Y	N	Y	N	N	N	Y	N	N	N	Y

Figure 6-1. Sample computer screens for test and study trials with 0 through 3 incorrect targets. *Note.* Allowable responses were limited to Y and N for test trials, and the number keys 1 through 5 for study trials. On the rating scale for study trials, 1 = not related, 3 = somewhat related, and 5 = highly related.

sequential fashion such that participants could not review previous responses. Participants were given up to ten minutes to complete the final test, and were stopped by the experimenter when this time has elapsed. Participants were then fully debriefed as to the purpose and expectations for the experiment. The entire procedure lasted approximately 45 minutes.

Results

Intervening Test Accuracy. The proportion of target items correctly recognized from the 0, 1, 2, and 3 incorrect target conditions was computed by calculating the proportion of hits minus false alarms for each target item. Intervening accuracy was very similar for targets preceded by 0 incorrect targets ($M = .95, SD = .10$), 1 incorrect target ($M = .96, SD = .10$), 2 incorrect targets ($M = .95, SD = .10$), and 3 incorrect targets ($M = .92, SD = .18$). These scores were analyzed using a one-way repeated measures ANOVA, which yielded a non-significant result [$F(3, 231) = 1.83, p > .05, MSE = .009$], indicating that intervening test accuracy did not vary significantly as a function of the number of incorrect targets presented.

Final Retention as a Function of Number of Incorrect Targets. The proportion of target items correctly recalled on the final test that were presented with 0 through 3 incorrect targets was computed for both TT and ST (see Figure 6-2 on the next page). These scores were analyzed using a 2 x 4 (Trial Type x Number of Targets) repeated-measures ANOVA. Significant main effects were obtained for trial type [$F(1, 77) = 31.64, p < .05, MSE = .02, \text{partial } \eta^2 = .29$], and for number of targets [$F(3, 231) = 4.12, p < .05, MSE = .02, \text{partial } \eta^2 = .05$]. From Figure 6-2, it is clear that tested items were retained significantly better than studied items, and retention of target items increased

when more incorrect targets were presented. Bonferroni's post-hoc tests indicated that targets preceded by 3 incorrect targets were retained significantly better than targets preceded by 0 incorrect targets, with no significant differences between any of the other pairwise comparisons. The Trial Type x Number of Targets Interaction was non-significant [$F(3, 231) = 1.20, p > .05, MSE = .02$], indicating that retention for tested versus studied items was not affected differently according to the number of incorrect targets presented.

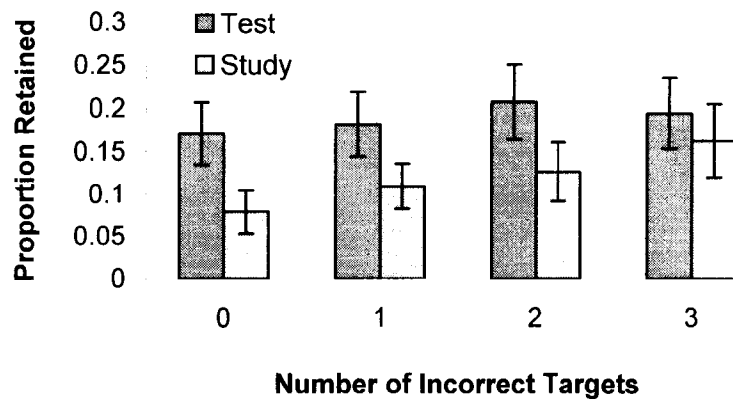


Figure 6-2. Proportion of items retained on final test as a function of type of trial and number of incorrect targets. Note. Error bars represent 95% confidence intervals.

Simple main effects analyses were carried out for both number of incorrect targets and trial. First, the effects of number of incorrect targets were analyzed separately for tested and studied items using one-way repeated-measures ANOVAs. These analyses revealed no significant effect of number of incorrect targets for tested items [$F(3, 231) = .71, p > .05, MSE = .03$], but did reveal a significant effect of number of incorrect targets for studied items [$F(2, 231) = 5.94, p < .05, MSE = .02, \text{partial } \eta^2 = .07$], and

Bonferroni's post-hoc tests revealed that the only significant pair-wise difference occurred between studied items preceded by 3 as opposed to 0 incorrect targets. Thus, it seems that the main effect for number of incorrect targets was driven by an increase in retention for items preceded by more as opposed to fewer incorrect targets, but only for studied items and not for tested items.

Second, the effects of type of trial were analyzed separately for 0 through 3 incorrect targets using paired-samples t-tests in which the alpha level was adjusted according to Bonferroni's correction across four comparisons ($.05 / 4 = .012$). These analyses revealed a significant testing effect for 0 incorrect targets [$t(77) = 4.52, p < .012, d = .51$], 1 incorrect target [$t(77) = 3.43, p < .012, d = .39$], and 2 incorrect targets [$t(77) = 3.17, p < .012, d = .36$], and a non-significant testing effect for 3 incorrect targets [$t(77) = 1.14, p > .012, d = .13$]. Thus, it seems that the main effect for type of trial was driven by a significant advantage for tested over studied items at all of the incorrect targets except for 3, and the strength of the testing effect decreased as the number of incorrect targets increased.

Discussion

The results of the simple main effects analyses for Experiment 3 confirm the predictions that retention for studied items would benefit significantly from the presentation of multiple incorrect targets, whereas retention for tested items would be unaffected by the presentation of multiple incorrect targets. Furthermore, the relative advantage for tested over studied items was strongest at 0 incorrect targets, and this relative advantage became weaker with increasing numbers of incorrect targets, until there was no significant advantage in retention for tested over studied items when 3

incorrect targets were presented. The results of these simple main effects analyses represent a trend in the direction of effects predicted by the multiple cue hypothesis.

However, lack of a significant Trial x Number of Targets Interaction provides no compelling evidence in favor of the multiple cue hypothesis to account for the testing effect. The significant main effect for type of trial, combined with the absence of a Trial x Number of Targets Interaction, suggests that tested items were retained significantly better than studied items, and retention for tested and studied items was not significantly different as a function of the number of incorrect targets provided. A significant interaction might be obtained if retention for tested items was more directly comparable at each level of incorrect target (rather than increasingly slightly from 0 to 2 incorrect targets, as shown in Figure 6-2), and if retention for studied items increased more sharply as a function of the number of incorrect targets.

There is reason to suspect that such an interaction did not occur because of the combination of two aspects of the wickelcall task used in Experiment 3. First, participants may have covertly recalled the correct target prior to making the recognition judgments, and second, the weak associations between cue-target pairs may have created conditions under which the presentation of multiple incorrect targets did not represent redundant processing for tested items. With regard to the early recall of correct targets, such a strategy (which was reported by one participant) would facilitate performance on test trials with an unpredictable number of incorrect targets, because participants would know ahead of time which target to look for, and thus be able to make the recognition judgments more efficiently. On all intervening test and study trials, the cue was presented alone for 1000 ms prior to the presentation of the target(s), so this may have allowed

enough time to retrieve the target or to begin searching memory to retrieve the target during the test trials. This would not be likely to happen during study trials, because participants were not required to make a recognition decision.

It is possible that the very nature of a serial yes/no recognition task involves recall processes (e.g., see Diller et al., 1998). Because of the fact that all of the potential targets are not presented at the same time, participants may have to recall the correct target in order to determine whether each option is correct or incorrect. Because of the fact that only one potential target was presented at a time, a participant must know what the correct target is, in order to make an accurate yes/no judgment when presented with multiple incorrect targets. This evidence suggests that it may be difficult to control the amount of extra information provided during a serial yes/no recognition task, because it is likely that recall processes are inherent in such a task. The amount of information sampled from memory during these recall processes is difficult to measure, so it is not clear whether similar amounts of extra information were provided on the intervening test and study trials of Experiment 3.

When using weakly related related cue-target pairs, it seems likely that the correct target will *not* be among the first information sampled from the cue (see Introduction to Experiment 2). If participants were indeed recalling the correct target prior to processing the multiple incorrect targets, and if participants' covert recall processes involved sampling information related to the cue, then the extra information provided by incorrect targets that were *unrelated* to the cues would not represent redundant processing for tested items. Instead, the combination of information sampled via recall and information sampled via presentation of incorrect targets represents a non-redundant convergence of

unrelated plus related information processed during test trials. This might account for the significant advantage of tested over studied items, as well as the slight increase in retention for tested items as a function of multiple incorrect targets.

The use of unrelated cue-target pairs may have also decreased the likelihood that incorrect targets would become associated with correct targets to a sufficient degree to serve as retrieval cues at the time of the final test. Thus, the unrelated associations between cue-target pairs may have played a role in the significant yet weak effect of the number of incorrect targets on retention for studied items (with an effect size of only .07). Perhaps the effectiveness of multiple cues would be increased if they were more related to target material, making it more likely that the number of cues would have a stronger positive effect on final retention.

The modified wickelcall task and the use of unrelated cue-target pairs in Experiment 3 did not insure that participants would process redundant information on test trials, nor did it insure that multiple cues would become adequately associated with targets to act as future retrieval cues. In order to adequately test the multiple cue hypothesis, an intervening memory test must be used in which the extra information provided must be redundant with what is ordinarily sampled during test trials but not study trials, and related enough to the target to act as future retrieval cues.

These issues could be addressed by replicating Experiment 3 using *related* as opposed to unrelated cue-target pairs. If information related to a cue is more likely to be sampled relative to information that is unrelated to a cue, then providing extra information that is related to the cue would be more likely to invoke redundant processing on test trials but not on study trials. Furthermore, related as opposed to unrelated cue-

target pairs would be more likely to become strongly associated with target information and serve as effective retrieval cues. Together, these effects might yield the significant Trial x Number of Targets Interaction predicted for Experiment 3, in which retention for tested items is comparable at each level of incorrect target, and retention for studied items increases strongly at each level of incorrect target.

CHAPTER VII

GENERAL DISCUSSION

Summary of Experiments

The results of all three experiments yielded several findings of theoretical importance concerning the testing effect. Experiment 1 addressed the question of whether or not tested items benefit from variable processing by presenting related cue-target pairs (*dough: bread*), followed by either a cued-recall test or an additional study opportunity in which the same cues were used that were encountered during the original encoding task (for test, *dough: _____*; for study, *dough: bread*), or different cues were used that were not encountered during the original encoding task (for test, *toast: _____*; for study, *toast: bread*). Results of Experiment 1 revealed a significant testing effect, and analysis of conditionalized scores revealed that both tested and studied items benefited significantly from different as opposed to the same cues.

Experiment 2 addressed the question of whether or not the strength of the testing effect is moderated by the degree of association between cue-target pairs. Based on two-process models of retrieval (Raaijmakers & Shiffrin, 1981) and Soraci et al.'s (1994, 1999) multiple cue hypothesis, remote cue-target associations as opposed to strong cue-target associations were expected to increase the likelihood that extra information would be sampled that may act as later retrieval cues, but only during a cued-recall test and not during an additional study opportunity. Experiment 2 presented either related (*dough: bread*) or unrelated (*basket: bread*) cue-target pairs, followed by an intervening cued-

recall test or additional study opportunity as in Experiment 1. Results of Experiment 2 revealed a significant testing effect, as well as a significant Trial x Cue Interaction in which unrelated cues significantly benefited retention for tested items but had no effect on retention for studied items.

Experiment 3 addressed the question of whether or not the amount of extra information associated with target items is directly related to retention of those target items. Unrelated cue-target pairs were presented, followed by an intervening yes/no recognition test or additional study opportunity. In the recognition test, the cue alone was presented briefly followed by the correct target after 0, 1, 2, or 3 incorrect targets, and in the study condition, each cue-target pair was rated for relatedness after rating the relatedness of the cue to 0, 1, 2, or 3 incorrect targets. Results of Experiment 3 revealed a significant testing effect, which decreased in magnitude as the number of incorrect targets increased, as well as a significant advantage for studied but not tested items that were presented with more as opposed to fewer incorrect targets. Although this pattern of simple main effects supported the predictions based on the multiple cue hypothesis, the lack of a significant Trial x Number of Targets Interaction precludes any strong conclusions in favor of this account of the testing effect.

Two factors that may have reduced the potential interaction effect in Experiment 3 are the tendency for participants to recall the correct target prior to processing the incorrect targets, and the tendency for unrelated cue-target pairs to serve as relatively weak future retrieval cues. Participants' reports suggested that the retention advantage for tested items may have been due to immediate recall processes involved in the modified wickelcall task, which created conditions under which non-redundant information—

specifically, information related to the cue, and information not related to the cue—was sampled on test trials, thus giving tested items a consistent advantage over studied items, as well as a modest and non-significant improvement in retention for tested items with increasing numbers of incorrect targets. The use of unrelated cue-target pairs not only made it possible to sample non-redundant information on test trials, but also made it possible that the extra information provided would not be effective in serving as retrieval cues later on, which may have accounted for the rather weak main effect of incorrect targets on studied items.

Memory for Tested Versus Studied Items

The results of all three experiments revealed a strong and reliable testing effect for the same target items across different experimental manipulations, including same versus different cues (Experiment 1), related versus unrelated cues (Experiment 2), and unrelated cues preceded by a number of incorrect targets (Experiment 3). These results add support to a large body of literature showing that tested items are better retained than studied items (e.g., Allen et al., 1969; Carrier & Pashler, 1992; Izawa, 1992; Kuo & Hirshman, 1996, 1997). Furthermore, the current study helps to rule out artifactual explanations for the testing effect based on the amount of processing or rehearsal devoted to tested items at the expense of studied items, by requiring active processing of studied items and manipulating test versus study trials between-lists.

The reliable testing effect obtained in the current experiments is consistent with a number of prior studies that have also used paired-associate learning in tests with feedback (Carrier & Pashler, 1992; Izawa, 1992), and in tests without feedback (Allen et al., 1969; Carpenter & DeLosh, in press; Cull, 2000). These past and current findings

suggest that the use of such a paired-associate learning paradigm is advantageous to testing effect research due to strong and reliable testing effects that are obtained, and also due to the degree of control that researchers have over cue and target characteristics (as opposed to, for example, free-recall tests) that may be important to the precise experimental manipulations designed to examine theoretical questions about the testing effect.

The specific pattern of results from Experiment 1 is consistent with past research investigating retention as a function of intervening cued-recall test trials in which the cues provided from encoding to the intervening test are either similar or different. Cuddy and Jacoby (1982) used a cued-recall test in which the orthographic letter cues provided during the intervening test trial were either the same or different from those provided during the first presentation. An example of the same cue would be *Laywer: C _ _ rt* and *Lawyer: C _ _ rt* during presentation and intervening test, respectively, whereas an example of a different cue would be *Laywer: C _ _ rt* and *Lawyer: _ ou _ _* during presentation and intervening test, respectively. Final cued-recall performance was greater for items that were given a different cue from presentation to intervening test, compared to items that were given the same cue from presentation to intervening test. The present study supplies much-needed data on variable processing and tested items, and replicates Cuddy and Jacoby's general pattern of results to show that retention for tested items can also benefit from the association of extra information provided by two cues as opposed to one.

There are currently no known studies of the testing effect that have directly compared retention for items tested with related as opposed to unrelated cues. so it is

difficult to determine whether the results of Experiment 2 are consistent with past research on the testing effect. These results are consistent, however, with research on memory phenomena similar to the testing effect, such as the generation effect, in which unrelated as opposed to related cues provide an advantage for retention. For example, Soraci et al. (1994) explored the effect of related versus unrelated cues in the generation effect by providing participants with a cue (e.g., *table*) and requiring them to read or generate an item related to the cue (e.g., *chair*), or an item unrelated to the cue (e.g., *tractor*). Final retention was better for generated items that were unrelated to the cue, compared to generated items that were related to the cue. Soraci et al. explained this effect by proposing that the generation of an unrelated cue, relative to a related cue, involves the generation of extra items that may act as cues for the correct item at the time of a later test. Experiment 2 demonstrated that this same pattern applies to the testing effect.

The results of Experiment 2 are also supported by unpublished data from our lab that show enhanced retention for tested items under conditions of impoverished cue support (Carpenter & DeLosh, accepted pending revisions). We used an orthographic cued-recall task in which participants were required to retrieve a previously-presented item using only one letter cue (*s _ _ _ _*), two letter cues (*s t _ _ _*), three letter cues (*s t r _ _*), or four letter cues (*s t r e _*). Similar to Experiment 2, this study created conditions under which access to the correct target (*street*) from the cue was more direct (in the case of four letter cues) or more remote (in the case of only one letter cue). Final retention for tested items was significantly higher for items tested with one letter cue as opposed to four letter cues. These results may be explained by the sampling of extra

information in testing conditions with more remote cue-target associations, as opposed to more direct cue-target associations. For example, several items other than *street* come to mind when presented with a single letter cue (*s _ _ _ _ _*), whereas few items other than *street* come to mind when presented with a four letter cue (*s t r e _ _*). Such results are consistent with the multiple cue hypothesis, and suggest that the extra information sampled during retrieval is not necessarily limited to items that are semantically associated to the targets, but may also include information that is orthographically related to the targets.

The results of Experiment 3 differ from those of a past study on the testing effect using a similar paradigm. Whitten and Leonard (1980) used a similar type of sequential yes/no recognition task and manipulated the number of distracters (2, 4, or 8) presented. Final free-recall retention increased significantly when more as opposed to fewer distracters were provided on the intervening recognition test. Thus, retention was enhanced by an intervening recognition test that was more difficult due to related distracters. Similarly, prior studies have also shown that retention is enhanced by intervening test conditions that promote interference from related material (Battig, 1979). This is presumably due to the additional or elaborative processing that occurs during difficult intervening test conditions, and that promotes resistance to forgetting over time (Glover, 1989; Gotz & Jacoby, 1978). Specifically, Whitten and Leonard propose that factors increasing the difficulty of intervening recognition tests (such as an increased number of distracters, or increased relatedness of distracters to targets) induce “a more extensive or more careful judgment of the target word’s familiarity, which, in turn, increases the probability that the target word will be recallable at some later time” (p.

130). In Experiment 3, final free-recall retention was not significantly affected by the number of distracters (i.e., incorrect targets) that were provided on the intervening recognition test, so it seems that these results do not support the *extended evaluation hypothesis* (Whitten & Leonard, 1980).

In order to investigate an explanation for the testing effect based on general difficulty or elaborative processing, a follow-up study to Experiment 3 was conducted in which retention for tested versus studied items was compared using only the 0 incorrect target condition. Experiment 3 was modified to present only 1 target, which was either correct or incorrect, in response to the cue. This way, participants were required to make only one judgment for each target item, representing a standard yes/no recognition task and eliminating the unpredictability of the number of incorrect targets. The new experimental design and counterbalancing versions for the follow-up study are presented in Appendix 6.

If the testing effect in Experiment 3 was driven by the unpredictable number of incorrect targets on test trials that may have induced a more extensive evaluation in order to identify the correct target, then no testing effect would be expected in a standard yes/no recognition task in which only one target is always presented. Contrary to this expectation, the follow-up experiment based on 22 new participants did yield a significant advantage in retention for tested items ($M = .17, SD = .09$) over studied items ($M = .11, SD = .08$), [$t(21) = 2.81, p < .05, d = .60$], and the proportions of tested and studied items retained were very similar to those of the overall proportions of tested ($M = .19$) and studied ($M = .12$) items retained from Experiment 3. Thus, it seems that the extended evaluation hypothesis does not fully account for the testing effect in Experiment

3. Instead, it is believed that the significant testing effect in Experiment 3 was due to the sampling of extra, non-redundant information during test trials relative to study trials.

Variable Processing in Memory

All three experiments also demonstrated reliable positive effects of variable processing. Consistent with past research on the beneficial effects of variable processing on studied items (Madigan, 1969; Young & Belleza, 1982), the current study demonstrated an improvement in retention for studied items that were associated with two different cues as opposed to the same cue (Experiment 1), and an improvement in retention for studied items that were associated with more as opposed to fewer distracters (Experiment 3). These results suggest that a greater amount of extra information associated with studied items enhances retention of those items, perhaps by increasing the number of potentially effective retrieval cues to those items. The role of variable processing in retention for tested items was less clear prior to the current study, as there was only one known study that examined retention for tested items using same versus different cues (Cuddy & Jacoby, 1982). The results of the current study therefore provide much-needed information regarding variable processing for tested items, and these results replicate and extend on the results of Cuddy and Jacoby by demonstrating an advantage in retention for tested items presented with two different cues as opposed to the same cue (Experiment 1), and an advantage in retention for tested items presented with unrelated cues as opposed to related cues (Experiment 2).

However, the mere *presentation* of extra information via same versus different cues is not enough to account for the testing effect. In Experiment 1, tested items that were given the same cue (for unconditionalized scores $M = .30$, $SD = .15$; for

conditionalized scores $M = .32$, $SD = .16$) were retained significantly better than studied items that were given a different cue ($M = .20$, $SD = .15$), despite the fact that extra information in the form of a different cue was presented in the latter relative to the former. This suggests that the mechanisms operating during retrieval in a cued-recall test contribute to better long-term retention than the presentation of additional information. A likely explanation is that extra information is routinely sampled on test trials prior to or in addition to the correct information, and this extra information acts as future retrieval cues for that correct information later on. If such extra information is routinely sampled on test trials but not on study trials, then the advantage of tested over studied items should occur even under conditions that do not promote increased sampling of extra information (such as the same cue condition of Experiment 1, and the related cue condition of Experiment 2), and this relative advantage should increase under conditions that promote increased sampling of extra information (such as the different cue condition of Experiment 1, and the unrelated condition of Experiment 2). This pattern was observed in both Experiments.

Experiment 3 set out to directly test the effects of sampling multiple targets during test as opposed to study. If multiple targets of a type similar to those presented in Experiment 3 are routinely sampled during test trials but not study trials, and these multiple targets are beneficial to retention, then the number of incorrect targets was expected to have no effect on final retention for tested items, but a positive effect on final retention for studied items. Although the data display a trend in this direction, the lack of the predicted interaction was confounded by the tendency for participants to recall the target items immediately and the use of unrelated cue-target pairs, both of which rendered the use of multiple incorrect targets non-redundant for tested items, and relatively weak

for studied items. An immediate direction for future research on this topic is to explore Experiment 3 further using related cue-target pairs, which may be more likely to represent the type of information routinely sampled during retrieval.

Past Research on the Testing Effect and Variable Processing

Although the testing effect has not received much theoretical attention from a variable processing perspective, the few studies that do exist suggest that tested items may be better retained than non-tested items due to enhanced variable processing as a result of intervening test trials. McDaniel and Masson (1985) found that test trials were more beneficial to memory retention when a different type of cue was given during presentation of cue-target pairs and the final cued-recall test (semantic followed by phonemic, or phonemic followed by semantic), rather than when the same type of cue was given during presentation and the final test (semantic followed by semantic, or phonemic followed by phonemic). They attributed this finding to the notion that the intervening cued-recall test increased the variability of the target item's memory trace, making it more likely to be retrieved in the context of a different cue rather than the same cue at the time of the final test. Such a retention advantage did not occur for intervening tests of recognition, which was interpreted to suggest that recognition tests do not increase the variable processing of tested items in the same way that intervening cued-recall tests do. Consistent with this is the tendency for forced-choice recognition tests to not benefit retention beyond a no-test control condition (Carpenter & DeLosh, accepted pending revisions; Glover, 1989). Rather than variable processing being induced by a specific type of test, however, the potential for variable processing may be a function of the degree to which an intervening test promotes recall processes that make it possible to

sample extra information from memory. As the current results demonstrate, this can indeed be accomplished with intervening recognition tests of the type used in Experiment 3.

It is also possible that tests benefit memory because the act of retrieval requires a convergence of both relational and item-specific processing, which is not required by a no-test control condition. Such a notion is derived from the literature on the dual benefits of organizational and distinctive processing in memory (Hunt & McDaniel, 1993), which proposes that memory benefits to the degree that organizational (relational information between items) and distinctive (item-specific information not shared by other items) processing converge upon the same item, and memory does not benefit when additional processing of only one type (relational or item-specific, but not both) is applied to the same item. Hunt and McDaniel illustrated the importance of processing relational and item-specific information by showing that memory for related word pairs (in which relational information is inherent) improved when the pairs were judged for differences (item-specific processing) rather than similarities (additional relational processing), and memory for unrelated word pairs (in which item-specific information is inherent) improved when the pairs were judged for similarities (relational processing) rather than differences (additional item-specific processing).

The role of relational and item-specific processing in retrieval was investigated by Matthews, Smith, Hunt, and Pivetta (1999), who required participants to remember items that were inherently related (grouped according to natural categories) or item-specific (grouped according to ad-hoc categories in which there was no apparent relationship), by performing a category sorting task (relational processing) or a pleasantness rating task

(item-specific processing). Participants then performed two types of memory tasks: a category cued-recall test in which they were required to retrieve the items from the categories when given the category label (which required episodic retrieval of the list items), and a category production test in which they were required to think of as many examples as they could from certain categories without regard to whether or not the items were from the lists (which did not require episodic retrieval of the list items). Matthews et al. found that conditions promoting the convergence of both relational and item-specific processing (pleasantness ratings for items in natural categories, or category ratings for items in ad-hoc categories) affected performance on the test of category cued-recall, whereas conditions promoting relational processing (natural categories and category ratings) but *not* item-specific processing (ad-hoc categories and pleasantness ratings) affected performance on the test of category production. Matthews et al. proposed that the act of retrieval requires relational information to organize the retrieval search, plus item-specific information to identify the correct items within that search. A non-retrieval-based task such as category production, on the other hand, does not require this convergence.

The data from Matthews et al. (1999) suggest that the act of retrieval requires relational processing to reinstate the organization of previously-encoded items, and item-specific processing to sample specific items and verify their accuracy. This reasoning is very similar to the two-process models of retrieval described in the Introduction, which forms the basis for the multiple cue hypothesis of the testing effect. Rather than a greater *amount* of information sampled during retrieval as opposed to a non-retrieval based task (such as additional presentation), Matthews et al. suggest that the information sampled during retrieval is more *unique* and less redundant compared to information sampled

during a non-retrieval-based task. This unique convergence of organizational and distinctive information has been shown to enhance memory in a number of studies (Hunt & McDaniel, 1996; McDaniel, Moore, & Whiteman, 1998).

The present set of experiments set out to further investigate the role of variable processing in the testing effect from a multiple cue perspective, and seem to provide evidence consistent with these past studies that intervening memory tests are beneficial to retention because they induce variable processing. It is not clear at this point whether the information processing associated with test trials is quantitative (in the form of multiple cues) or qualitative (in the form of non-redundant processing), or perhaps some combination of both. In order to explore this question, future follow-up research could include the use of relation *and* item-specific processing in tasks that are similar to the ones used in this study, for example, requiring ratings of relatedness as well as differences between items rather than a dual-dose of relatedness for study trials. Past and current data on this topic suggest that exploring the qualitative and quantitative nature of information associated with test and study conditions represents a viable area for further exploration.

Comparing the Elaborative Retrieval and Multiple Cue Hypotheses

The question arises as to the degree to which the current data uniquely support the multiple cue hypothesis beyond the more general elaborative retrieval hypothesis. The elaborative retrieval hypothesis proposes that more difficult retrieval conditions as opposed to less difficult retrieval conditions invoke additional or more extensive processing, which accounts for enhanced retention later on. Therefore, any manipulations increasing the difficulty of the intervening test retrieval conditions should be associated with better final test retention. Intervening test retrieval rates indicate that across the three

experiments, retrieval was more difficult for the different as opposed to same cue conditions (Experiment 1), for the unrelated as opposed to related cue conditions (Experiment 2), and did not differ across the range of number of incorrect targets presented (Experiment 3). Consistent with the predictions of the elaborative processing hypothesis, final retention was better for items tested with different as opposed to the same cues (Experiment 1), for unrelated as opposed to related cues (Experiment 2), and did not differ across the range of number of incorrect targets presented (Experiment 3). Is there evidence that uniquely supports the multiple cue hypothesis, in which extra information sampled during retrieval accounts for the testing effect, over and beyond the elaborative retrieval hypothesis, in which effortful or extensive processing accounts for the testing effect?

Some of the data provided in this study, particularly those of the follow-up study to Experiment 3, can provide an answer this question. The follow-up study to Experiment 3 examined the extensive evaluation hypothesis (Whitten & Leonard, 1980) to determine whether the observed testing effect was due to general difficulty associated with an unpredictable number of distracters on the recognition test. The recognition test in the follow-up study was made easier, relative to the recognition test in Experiment 3, by only providing one item to be recognized on all trials (the intervening retrieval rates for the 0 incorrect target condition in Experiment 3 and the follow-up study were .95 and .99, respectively). This manipulation should have reduced the degree of extended evaluations taking place on the recognition test. Despite this, a significant testing effect still emerged, and the retention rates for tested and studied items in the follow-up study were highly similar to those in Experiment 3. These results suggest that the degree of extended

evaluation, manipulated by the number of distracters, did not play a major role in the testing effect observed in these experiments.

The elaborative processing hypothesis is difficult to rule-out in all situations because of its generality. However, the current data offer a strong argument in favor of a more precise version of the elaborative processing hypothesis to account for the testing effect, specifically, one based on the multiple cue hypothesis. This represents an important step toward presenting a novel explanation of the testing effect that is based on well-known memory principles, that is well-specified, and most importantly, that is testable and falsifiable. The current study not only serves to provide a starting point for a broad range of future studies seeking to further test the multiple cue hypothesis for the testing effect, but also provides data suggesting that future investigations on this topic are necessary and worthwhile.

References

- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463-470.
- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97-123.
- Bahrick, H. A. (1970). Two-phase model for prompted recall. *Psychological Review*, 77, 215-222.
- Bartlett, J. C. (1977). Effects of immediate testing and delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 719-732.
- Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory* (pp. 23-44). Mahwah, NJ: Erlbaum.
- Birnbaum, I. M., & Eichner, J. (1971). Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 10, 516-521.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 397-401). New York: Academic Press.

- Bjork, R. A., & Whitten, W. B. (1974). Recency sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*, 173-189.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*, (pp. 85-123). Washington, D. C.: V. H. Winston.
- Brainerd, C. J., Wright, R., & Reyna, V. F. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language*, *46*, 120-152.
- Bregman, A. S., & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *9*, 689-698.
- Bugelski, R. R. (1962). Presentation time, total time and mediation in paired associate learning. *Journal of Experimental Psychology*, *63*, 409-412.
- Carpenter, S. K., & DeLosh, E. L. (accepted pending revisions). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*.
- Carpenter, S. K., & DeLosh, E. L. (in press). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, *20*, 633-642.
- Cooper, E. H., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin*, *68*, 221-234.
- Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*, 451-467.

- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215-235.
- Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding the understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2*, 365-378.
- Darley, D. F., & Murdock, B. B. Jr. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 93*, 66-73.
- Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 414-435.
- Donaldson, W. (1971). Output effects in multitrial free-recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 577-585.
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*, 94-107.
- Estes, W. K. (1955a). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.
- Estes, W. K. (1955b). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145-154.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*, 392-399.
- Glover, J. A., Krug, D., Hannon, S., & Shine, A. (1990). The “testing” effect and restricted retrieval rehearsal. *The Psychological Record, 40*, 215-226.

- Gotz, A., & Jacoby, L. L. (1974). Encoding and retrieval processes in long-term retention. *Journal of Experimental Psychology, 2*, 291-297.
- Green, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371-377.
- Green, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1004-1011.
- Green, R. L., & Stillwell, A. M. (1995). Effects of encoding variability and spacing on frequency discrimination. *Journal of Memory and Language, 34*, 468-476.
- Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47-91). New York: Academic Press.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.
- Izawa, C. (1992). Test trial contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy & S. M. Kosslyn (Eds.), *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory* (pp. 1-33). Hillsdale, NJ: Erlbaum.
- Kuo, T. M., & Hirshman, E. (1997). The role of distinctive perceptual information in memory: Studies of the testing effect. *Journal of Memory and Language, 36*, 188-201.
- Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109*, 451-464.

- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.
- Landauer, T. K., & Eldridge, L. (1967). Effects of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology*, *75*, 290-298.
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828-835.
- Matthews, T. D., Smith, R. E., Hunt, R. R., & Pivetta, C. E. (1999). Role of distinctive processing during retrieval. *Psychological Reports*, *84*, 904-916.
- McDaniel, M. A., & Fisher, R. P. (1991) Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192-201.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 371-385.
- McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 173-185.
- McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory and Cognition*, *17*, 423-434.

- Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 609-622.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519-533.
- Morris, P. E., Fritz, C. O., & Buck, S. (2004). The name game: Acceptability, bonus information and group size. *Applied Cognitive Psychology*, 18, 89-104.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*.
<http://www.usf.edu/FreeAssociation/>
- Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 384-413.
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, 74, 18-22.
- Petros, T., & Hoving, K. (1980). The effects of review on young children's memory for prose. *Journal of Experimental Child Psychology*, 30, 33-43.
- Postman, L., & Phillips, L. W. (1961). Studies in incidental learning: A comparison of the methods of successive and single recalls. *Journal of Experimental Psychology*, 61, 236-241.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.

- Rothkopf, E. Z., & Coke, E. U. (1963). Repetition interval and rehearsal method in learning equivalences from written sequences. *Journal of Verbal Learning and Verbal Behavior*, 2, 406-416.
- Rothkopf, E. Z., & Coke, E. U. (1966). The effect of some variations in procedure on response repetition following verbal outcomes. *Journal of Verbal Learning and Verbal Behavior*, 5, 86-91.
- Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory and Cognition*, 11, 641-650.
- Runquist, W. N. (1986). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, 40, 65-76.
- Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, 26, 589-607.
- Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 716-727.
- Soraci, S. A., Jr., Carlin, M. T., Chechile, R. A., Franks, J. J., Wills, T., & Watanabe, T. (1999). Encoding variability and cueing in generative processing. *Journal of Memory and Language*, 41, 541-559.
- Soraci, S. A., Jr., Franks, J. J., Bransford, J. D., Chechile, R. A., Belli, R. F., Carr, M., & Carlin, M. (1994). Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 67-78.

- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175-184.
- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 135-144.
- Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240-245.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16, 465-478.
- Whitten, W. B., & Leonard, J. M. (1980). Learning from tests: Facilitation of delayed recall by initial recognition alternatives. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 127-134.
- Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 189-202.
- Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments, and Computers*, 20, 6-11.
- Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 545-559.

APPENDIX I

Targets, cues, and cue-to-target strength for stimuli used in Experiments 1 and 2

Target	Experiment 1				Experiment 2			
	Cue 1		Cue 2		Related Cue		Unrelated Cue	
Bread	Dough	(.31)	Toast	(.36)	Toast	(.36)	Basket	(.01)
Court	Trial	(.25)	Jury	(.25)	Jury	(.25)	Rights	(.01)
Heart	Transplant	(.40)	Valentine	(.42)	Valentine	(.42)	Rib	(.01)
Horse	Stable	(.42)	Rodeo	(.47)	Rodeo	(.47)	Camel	(.01)
Music	Symphony	(.32)	Stereo	(.33)	Stereo	(.33)	Theater	(.01)
Planet	Venus	(.37)	Neptune	(.39)	Neptune	(.39)	Comet	(.01)
Smoke	Puff	(.24)	Chimney	(.24)	Chimney	(.24)	Fire	(.01)
Teeth	Orthodontist	(.41)	Dentist	(.45)	Dentist	(.45)	Lips	(.01)
Block	Brick	(.50)	Chunk	(.50)	Chunk	(.50)	Chisel	(.01)
Child	Parent	(.23)	Adolescent	(.26)	Adolescent	(.26)	Mitten	(.01)
Coffee	Café	(.32)	Tea	(.36)	Tea	(.36)	Steam	(.01)
Doctor	Clinic	(.30)	Patient	(.36)	Patient	(.36)	Virus	(.01)
Grass	Green	(.25)	Mow	(.27)	Mow	(.27)	Picnic	(.01)
Knife	Blade	(.34)	Fork	(.37)	Fork	(.37)	Kitchen	(.01)
Money	Tax	(.41)	Cost	(.41)	Cost	(.41)	Contest	(.01)
Train	Conductor	(.28)	Station	(.31)	Station	(.31)	Airplane	(.01)
Blood	Vampire	(.33)	Vein	(.38)	Vein	(.38)	Bruise	(.01)
Circle	Diameter	(.25)	Sphere	(.25)	Sphere	(.25)	Hole	(.01)
Dress	Hem	(.26)	Skirt	(.29)	Skirt	(.29)	Maid	(.01)
Flower	Stem	(.43)	Blossom	(.44)	Blossom	(.44)	Leaf	(.01)
Hotel	Lobby	(.34)	Suite	(.35)	Suite	(.35)	Lounge	(.01)
Island	Paradise	(.27)	Adventure	(.29)	Adventure	(.29)	Desert	(.01)
Light	Beam	(.45)	Switch	(.45)	Switch	(.45)	Morning	(.01)
School	Principle	(.31)	Education	(.31)	Education	(.31)	Pupil	(.01)
Brush	Paint	(.32)	Bristle	(.39)	Bristle	(.39)	Flick	(.01)
Drink	Cocktail	(.45)	Beverage	(.49)	Beverage	(.49)	Ice	(.01)
Floor	Ceiling	(.21)	Mop	(.24)	Mop	(.24)	Elevator	(.01)
House	Mansion	(.32)	Home	(.33)	Home	(.33)	Barn	(.01)
Movie	Preview	(.25)	Video	(.25)	Video	(.25)	Television	(.01)
Paper	Margin	(.32)	Folder	(.32)	Folder	(.32)	Scissors	(.01)
River	Flow	(.28)	Stream	(.32)	Stream	(.32)	Bay	(.01)
Street	Sidewalk	(.28)	Main	(.31)	Main	(.31)	Directions	(.01)
Beach	Seashore	(.21)	Shell	(.25)	Shell	(.25)	Raft	(.01)
Chair	Desk	(.28)	Couch	(.29)	Couch	(.29)	Hammock	(.01)
Clock	Time	(.37)	Alarm	(.38)	Alarm	(.38)	Dial	(.01)
Dance	Tap	(.20)	Prom	(.22)	Prom	(.22)	Spin	(.01)
Forest	Woods	(.23)	Wilderness	(.26)	Wilderness	(.26)	Meadow	(.01)
Glass	Breakable	(.38)	Shatter	(.41)	Shatter	(.41)	Bead	(.01)
Mouth	Jaw	(.25)	Saliva	(.26)	Saliva	(.26)	Speak	(.01)
Stone	Granite	(.23)	Rock	(.26)	Rock	(.26)	Building	(.01)
Church	Prayer	(.34)	Minister	(.34)	Minister	(.34)	Soul	(.01)
Dinner	Banquet	(.26)	Lunch	(.26)	Lunch	(.26)	Manners	(.01)
Fence	Gate	(.35)	Picket	(.38)	Picket	(.38)	Barrier	(.01)
Fruit	Berry	(.29)	Plum	(.29)	Plum	(.29)	Seed	(.01)
Letter	Mailbox	(.27)	Note	(.29)	Note	(.29)	Print	(.01)
Phone	Operator	(.35)	Call	(.37)	Call	(.37)	Ear	(.01)
Shirt	Button	(.24)	Cuff	(.24)	Cuff	(.24)	Jacket	(.01)
Water	Mineral	(.47)	Hose	(.47)	Hose	(.47)	Mist	(.01)

Note: Cue-to-target strength is given in parentheses next to each cue.

APPENDIX II

Experiment versions for Experiments 1 and 2

Version 1a		Type of Cue	
List	Type of Trial	Experiment 1	Experiment 2
1	Test	Same	Related
2	Study	Different	Unrelated
3	Study	Same	Related
4	Test	Same	Related
5	Test	Different	Unrelated
6	Study	Different	Unrelated

Version 1b			
List	Type of Trial	Experiment 1	Experiment 2
1	Test	Different	Unrelated
2	Study	Same	Related
3	Study	Different	Unrelated
4	Test	Different	Unrelated
5	Test	Same	Related
6	Study	Same	Related

Version 2a			
List	Type of Trial	Experiment 1	Experiment 2
1	Study	Same	Related
2	Test	Different	Unrelated
3	Test	Same	Related
4	Study	Same	Related
5	Study	Different	Unrelated
6	Test	Different	Unrelated

Version 2b			
List	Type of Trial	Experiment 1	Experiment 2
1	Study	Different	Unrelated
2	Test	Same	Related
3	Test	Different	Unrelated
4	Study	Different	Unrelated
5	Study	Same	Related
6	Test	Same	Related

APPENDIX III

Alternative data analysis for Experiment 1, based on scores for 33 participants whose intervening test retrieval rate was greater than .66, and scores for 14 participants whose intervening test retrieval rate was greater than .75.

Table 4-2.

Intervening Retrieval Rate	Type of Trial	
	Test	Study
Greater than .66 ($N = 33$)		
Same Cue	.29 (.15)	.16 (.15)
Different Cue	.31 (.16)	.20 (.16)
Greater than .75 ($N = 14$)		
Same Cue	.31 (.15)	.20 (.14)
Different Cue	.37 (.17)	.25 (.17)

Note: Standard deviations are given in parentheses.

A re-analysis of the final test data using the 14 participants whose retrieval rates were the most comparable shows that the main effect for trial still emerges [$F(1, 13) = 24.34, p < .05, MSE = .007, \text{partial } \eta^2 = .65$] in favor of test trials, a marginally significant main effect for cue also emerges [$F(1, 13) = 3.62, p = .08, MSE = .012, \text{partial } \eta^2 = .22$] in favor of different cues, and the Trial x Cue Interaction is no longer significant [$F(1, 13) = .009, p > .05, MSE = 6.98$].

APPENDIX IV

Cues, targets, and cue-to-target strength for stimuli used in Experiment 3

Cue	Correct Target		Incorrect Targets					
Basket	Bread	(.01)	Carry	(.01)	Eggs	(.01)	Wicker	(.04)
Rights	Court	(.01)	Choice	(.01)	Legal	(.03)	People	(.02)
Rib	Heart	(.01)	Chest	(.04)	Lung	(.02)	Stomach	(.01)
Camel	Horse	(.01)	Back	(.01)	Donkey	(.01)	Egypt	(.02)
Theater	Music	(.01)	Actor	(.03)	Drama	(.04)	Stage	(.01)
Comet	Planet	(.01)	Meteor	(.04)	Sky	(.04)	Universe	(.01)
Fire	Smoke	(.01)	Flame	(.04)	Burn	(.05)	Engine	(.02)
Lips	Teeth	(.01)	Face	(.01)	Lipstick	(.01)	Tongue	(.03)
Chisel	Block	(.01)	Artist	(.01)	Tool	(.05)	Fork	(.01)
Mitten	Child	(.01)	Cat	(.01)	Wool	(.01)	Hand	(.05)
Steam	Coffee	(.01)	Bath	(.02)	Tea	(.01)	Kettle	(.04)
Virus	Doctor	(.01)	Fever	(.02)	Illness	(.02)	Germ	(.02)
Picnic	Grass	(.01)	Blanket	(.01)	Park	(.05)	Sunshine	(.01)
Kitchen	Knife	(.01)	Counter	(.02)	Dish	(.01)	Stove	(.05)
Contest	Money	(.01)	Compete	(.04)	Fight	(.01)	Sport	(.01)
Airplane	Train	(.01)	Boat	(.01)	Pilot	(.04)	Wings	(.03)
Bruise	Blood	(.01)	Scar	(.01)	Wound	(.01)	Ouch	(.01)
Hole	Circle	(.01)	Dark	(.02)	Round	(.02)	Pit	(.01)
Maid	Dress	(.01)	Milk	(.02)	Service	(.02)	Uniform	(.01)
Leaf	Flower	(.01)	Branch	(.02)	Fall	(.03)	Oak	(.01)
Lounge	Hotel	(.01)	Lizard	(.03)	Pool	(.01)	Room	(.01)
Desert	Island	(.01)	Cactus	(.02)	Hot	(.05)	Oasis	(.02)
Morning	Light	(.01)	Afternoon	(.04)	Sleep	(.02)	Star	(.02)
Pupil	School	(.01)	Eye	(.16)	Student	(.68)	Teacher	(.08)
Flick	Brush	(.01)	Switch	(.01)	Throw	(.02)	Flip	(.02)
Ice	Drink	(.01)	Bucket	(.01)	Pick	(.03)	Skate	(.02)
Elevator	Floor	(.01)	Box	(.02)	Button	(.01)	Door	(.02)
Barn	House	(.01)	Loft	(.02)	Shed	(.01)	Stable	(.01)
Television	Movie	(.01)	Channel	(.01)	Station	(.02)	Screen	(.01)
Scissors	Paper	(.01)	Sharp	(.02)	Cut	(.87)	Sew	NA
Bay	River	(.01)	Harbor	(.01)	Lake	(.01)	Shore	(.02)
Directions	Street	(.01)	Rules	(.03)	Way	(.03)	Test	(.01)
Raft	Beach	(.01)	Air	(.01)	Canoe	(.01)	Log	(.03)
Hammock	Chair	(.01)	Tree	(.04)	Nap	(.03)	Lazy	(.02)
Dial	Clock	(.01)	Knob	(.02)	Number	(.01)	Sun	(.01)
Spin	Dance	(.01)	Twist	(.01)	Weave	(.01)	Wheel	(.01)
Meadow	Forest	(.01)	Land	(.02)	Prairie	(.01)	Sheep	(.01)
Bead	Glass	(.01)	Chain	(.01)	Pearl	(.02)	Wax	(.01)
Speak	Mouth	(.01)	Hear	(.01)	Language	(.01)	Voice	(.02)
Building	Stone	(.01)	Concrete	(.02)	Office	(.04)	City	(.01)
Soul	Church	(.01)	Mind	(.02)	Body	(.03)	Self	(.02)
Manners	Dinner	(.01)	Eating	(.01)	Proper	(.02)	Morals	(.01)
Barrier	Fence	(.01)	Road	(.01)	Gate	(.01)	Line	(.01)
Seed	Fruit	(.01)	Bird	(.02)	Farm	(.01)	Weed	(.02)
Print	Letter	(.01)	Page	(.02)	Spell	(.02)	Words	(.02)
Ear	Phone	(.01)	Ache	(.03)	Listen	(.03)	Sound	(.04)
Jacket	Shirt	(.01)	Pants	(.03)	Suit	(.01)	Tie	(.03)
Mist	Water	(.01)	Cloud	(.02)	Haze	(.04)	Wet	(.01)

Note: Cue-to-target strength is given in parentheses next to each target.

APPENDIX V

Experimental versions for Experiment 3 (Versions are 1a, 1b, 1c, 1d, 2a, 2b, 2c, 2d)

VERSION 1a-1d					VERSION 2a-2d				
List 1 (TT)	Number of Extra Targets				List 1 (ST)	Number of Extra Targets			
	1a	1b	1c	1d		2a	2b	2c	2d
Bread	0	3	2	1	Bread	0	3	2	1
Court	1	0	3	2	Court	1	0	3	2
Heart	2	1	0	3	Heart	2	1	0	3
Horse	3	2	1	0	Horse	3	2	1	0
Music	0	3	2	1	Music	0	3	2	1
Planet	1	0	3	2	Planet	1	0	3	2
Smoke	2	1	0	3	Smoke	2	1	0	3
Teeth	3	2	1	0	Teeth	3	2	1	0
List 2 (ST)					List 2 (TT)				
Block	0	3	2	1	Block	0	3	2	1
Child	1	0	3	2	Child	1	0	3	2
Coffee	2	1	0	3	Coffee	2	1	0	3
Doctor	3	2	1	0	Doctor	3	2	1	0
Grass	0	3	2	1	Grass	0	3	2	1
Knife	1	0	3	2	Knife	1	0	3	2
Money	2	1	0	3	Money	2	1	0	3
Train	3	2	1	0	Train	3	2	1	0
List 3 (ST)					List 3 (TT)				
Blood	0	3	2	1	Blood	0	3	2	1
Circle	1	0	3	2	Circle	1	0	3	2
Dress	2	1	0	3	Dress	2	1	0	3
Flower	3	2	1	0	Flower	3	2	1	0
Hotel	0	3	2	1	Hotel	0	3	2	1
Island	1	0	3	2	Island	1	0	3	2
Light	2	1	0	3	Light	2	1	0	3
School	3	2	1	0	School	3	2	1	0
List 4 (TT)					List 4 (ST)				
Brush	0	3	2	1	Brush	0	3	2	1
Drink	1	0	3	2	Drink	1	0	3	2
Floor	2	1	0	3	Floor	2	1	0	3
House	3	2	1	0	House	3	2	1	0
Movie	0	3	2	1	Movie	0	3	2	1
Paper	1	0	3	2	Paper	1	0	3	2
River	2	1	0	3	River	2	1	0	3
Street	3	2	1	0	Street	3	2	1	0
List 5 (TT)					List 5 (ST)				
Beach	0	3	2	1	Beach	0	3	2	1
Chair	1	0	3	2	Chair	1	0	3	2
Clock	2	1	0	3	Clock	2	1	0	3
Dance	3	2	1	0	Dance	3	2	1	0
Forest	0	3	2	1	Forest	0	3	2	1
Glass	1	0	3	2	Glass	1	0	3	2
Mouth	2	1	0	3	Mouth	2	1	0	3
Stone	3	2	1	0	Stone	3	2	1	0
List 6 (ST)					List 6 (TT)				
Church	0	3	2	1	Church	0	3	2	1
Dinner	1	0	3	2	Dinner	1	0	3	2
Fence	2	1	0	3	Fence	2	1	0	3
Fruit	3	2	1	0	Fruit	3	2	1	0
Letter	0	3	2	1	Letter	0	3	2	1
Phone	1	0	3	2	Phone	1	0	3	2
Shirt	2	1	0	3	Shirt	2	1	0	3
Water	3	2	1	0	Water	3	2	1	0

Note. Test trials and study trials are denoted by TT and ST, respectively.

APPENDIX VI

Experimental versions for follow-up study to Experiment 3 (versions are 1a, 1b, 2a, 2b)

List 1	1a	1b	2a	2b	List 4	1a	1b	2a	2b
Bread	TTn	TTv	STn	STv	Brush	TTn	TTv	STn	STv
Court	TTn	TTv	STn	STv	Drink	TTn	TTv	STn	STv
Heart	TTn	TTv	STn	STv	Floor	TTn	TTv	STn	STv
Horse	TTn	TTv	STn	STv	House	TTn	TTv	STn	STv
Music	TTv	TTn	STv	STn	Movie	TTv	TTn	STv	STn
Planet	TTv	TTn	STv	STn	Paper	TTv	TTn	STv	STn
Smoke	TTv	TTn	STv	STn	River	TTv	TTn	STv	STn
Teeth	TTv	TTn	STv	STn	Street	TTv	TTn	STv	STn
List 2					List 5				
Block	STn	STv	TTn	TTv	Beach	TTn	TTv	STn	STv
Child	STn	STv	TTn	TTv	Chair	TTn	TTv	STn	STv
Coffee	STn	STv	TTn	TTv	Clock	TTn	TTv	STn	STv
Doctor	STn	STv	TTn	TTv	Dance	TTn	TTv	STn	STv
Grass	STv	STn	TTv	TTn	Forest	TTv	TTn	STv	STn
Knife	STv	STn	TTv	TTn	Glass	TTv	TTn	STv	STn
Money	STv	STn	TTv	TTn	Mouth	TTv	TTn	STv	STn
Train	STv	STn	TTv	TTn	Stone	TTv	TTn	STv	STn
List 3					List 6				
Blood	STn	STv	TTn	TTv	Church	STn	STv	TTn	TTv
Circle	STn	STv	TTn	TTv	Dinner	STn	STv	TTn	TTv
Dress	STn	STv	TTn	TTv	Fence	STn	STv	TTn	TTv
Flower	STn	STv	TTn	TTv	Fruit	STn	STv	TTn	TTv
Hotel	STv	STn	TTv	TTn	Letter	STv	STn	TTv	TTn
Island	STv	STn	TTv	TTn	Phone	STv	STn	TTv	TTn
Light	STv	STn	TTv	TTn	Shirt	STv	STn	TTv	TTn
School	STv	STn	TTv	TTn	Water	STv	STn	TTv	TTn

Note. Test trials and study trials are denoted by *TT* and *ST*, respectively, and cases in which the correct target was provided are denoted as *y*, whereas cases in which an incorrect target was provided are denoted as *n*. The incorrect targets were taken from the leftmost column of incorrect targets in Appendix 4.