Investigating Metacognitive Control in a Global Memory Framework

Kenneth J. Malmberg

University of South Florida

Corresponding Author:
Kenneth J. Malmberg
Department of Psychology
University of South Florida
4202 East Fowler Ave., PCD3118G
Tampa, FL 34630

email: malmberg@cas.usf.edu
phone: (813) 974-7641

How does one learn? How does one remember? These are the broad questions that the Nelson and Narens (1990) research program addressed. Of course, they were not the first to ask these questions, but they did approach these questions in a novel way.

The Nelson and Narens approach to understanding learning and memory can be viewed as an extension of Atkinson and Shiffrin's (1968) proposal that memory consists of a set of memory structures and control processes. The memory structures are assumed to be used to support the performance of all learning and memory tasks, whereas control processes (e.g., rehearsal) are assumed to be strategically used to perform particular tasks. Many researchers have sought to understand the nature of the structural aspects of learning memory, and this has led to several formal models. Nelson and Narens, on the other hand, organized the prevalent measures and developed a framework that describes how the structural aspects of memory are monitored and controlled. It is a testament to the empirical richness of the Nelson and Narens *metamemory* framework that those modern researchers who investigate metamemory do so largely independently of those who investigate the structural aspects of memory (and vice versa). In this chapter, I consider how these two approaches to understanding learning and memory might be jointly used to build better models of learning and memory.

Retrieval and Matching in Memory

Global theories of memory attempt to explain a large number of memory phenomena with just a few central assumptions. They often describe remembering as an interaction between retrieval cues and memory. That is, memory is queried by probing it with a set of information that represents the nominal stimulus and the result of the probe depends on the nature of the information in the retrieval cue. Typically it is assumed that memory traces are activated or accessible to the extent that they contain information that is similar to the contents of the retrieval cue and to the extent that they are well encoded.

Most theories of episodic memory propose that two types of processes access the information stored in memory (e.g., Gillund & Shiffrin, 1984; Hintzman, 1987; Humphreys, Bain, & Pike, 1989; Murdock, 1993; Shiffrin & Steyvers, 1997). I will refer to these as *retrieval* and *global-matching* processes, and they produce qualitatively different types of information (cf. Humphreys, et al., 1989). A retrieval process provides information about the contents of a memory trace, while a global-matching process provides information about the familiarity of a retrieval cue. The later process is referred to as “global-matching” because the retrieval cue is compared to the contents of large number (perhaps all) traces in memory. Thus, familiarity is assumed to be a positive function of the similarity between these memory traces and the retrieval cue.

For instance, let us assume that one has studied a pair of words – TROUT and PINT. If subsequently presented with TROUT, one might probe memory with the orthographic, phonologic, and semantic information associated with it. The probability of then retrieving PINT would be a positive function of how well encoded TROUT and PINT were during study. In addition, having been presented with TROUT one almost certainly would have some sense that it was recently encountered (i.e., it seems familiar) independently of the ability to retrieve PINT, and the longer TROUT was studied or the more times TROUT was studied the better encoded it would be and hence the more familiar it would seem.

Accordingly, free or cued recall tasks are generally assumed to involve a retrieval process, while recognition tasks are often assumed to involve a global-matching process (Gillund & Shiffrin, 1984; Hintzman, 1987; Humphreys et al., 1989; Malmberg, Zeelenberg, & Shiffrin, 2004; Murdock, 1993; Shiffrin & Steyvers, 1997). In some theories of recognition memory, output from the global-matching process (e.g., familiarity) serves as the input to a decision mechanism that is modeled by a version of signal-detection theory to produce a response. Other theories of recognition assume that recognition is based on the operation of both retrieval and a global-matching process (e.g., Atkinson & Juola, 1973; Malmberg, Holden, & Shiffrin, 2004; Reder et al., 2000; Mandler, 1980; see Clark, 1998 ; Mandler, 1991; Yonelinas, 2002 for reviews). A major topic of research has been to empirically test these two models of recognition. Less attention has been given to what role, if any, familiarity plays in free or cued recall, although I will discuss some relevant findings below. One reason for this comparative lack of interest by memory researchers is that familiarity alone is insufficient for successfully performing a recall task; recall demands a response that names an item and the matching process does not produce items as output. A second reason lays in the limited scope of many memory theories.

Search Permission and Familiarity

Memory control processes generally produce the input for the retrieval process, and they make use of the output from the retrieval process to govern the completion of a memory task. With several exceptions (e.g., Atkinson & Shiffrin, 1968; Malmberg & Xu, in press; Raaijmakers & Shiffrin, 1981), memory control processes have not been modeled in great detail. Consideration of a range of possible control processes provides a rich field of possibilities for the use of familiarity in recall. [1] For example, Diller, Nobel, and Shiffrin (2001) assume in their REM model of cued recall that the amount of time subjects are willing to search memory is positive function of the familiarity of the retrieval cue.

Does familiarity affect the amount of time one is willing to search memory in a cued recall task? Convergent empirical support for the hypothesis that the familiarity produced by the retrieval cue is used to control memory search comes from several investigations of metacognitive feeling-of-knowing judgments (Koriat, 1993; Metcalfe, 1993; Reder, 1987; Schwartz & Metcalfe,1992; also see Glucksberg & McCloskey, 1981). For instance, some have proposed that the length of a search is based on a chain of events beginning with memory access (Nelson & Narens, 1990; Reder, 1987). A feeling-of-knowing judgment is made when retrieval fails, and additional attempts to remember are likely when feeling-of-knowing judgments are positive (Nelson & Narens, 1990). Several investigators have proposed that feeling-of-knowing judgments are informed, at least in part, by the familiarity produced by the retrieval cue (Koriat, 1993; Nelson, Gerler, & Narens, 1984; Reder 1987; Metcalfe, 1993).

Schwartz and Metcalfe (1992) and Metcalfe, Schwartz, and Joaquin (1993) confirmed a straightforward prediction of this hypothesis: Directly priming a cue produces greater feeling-of-knowing judgments. Nelson et al. (1984) reported a positive correlation between feeling-of-knowing judgments and the length of a search for answers to general knowledge questions. Reder (1987) reported longer search times in response to primed normatively-difficult general knowledge questions but shorter search times in response to primed normatively-easy questions (Reder 1987, Exp. 6). Thus, there is some evidence that cue familiarity does inform the decision of when to terminate a search of semantic memory. It remains, however, an open question as to whether the familiarity of the retrieval cue affects the length of search for episodic memory tasks, like paired-associated cued recall, and whether there are any empirical limitations to such a model.

Hypotheses and Predictions

Here I report the results of four paired-associate cued recall experiments. Pairs of words were studied and one word was presented as a cue to recall the other word at test. The responses were divided into two categories for the present analyses: correct responses and "don't know" responses. The interests here are how cue familiarity affects the willingness to search memory (or length of search) and how this might affect recall performance. The first interest is inherently a metamemory issue and the latter is primarily a structural memory issue.

To address these issues, I will measure both the accuracy and the latency of cued recall performance. The latencies of correct responses do not provide a good indicator of maximum search time because a search may have continued longer if not for the retrieval of an item deemed worthy of reporting (cf. Gillund & Shiffrin, 1984; Nelson & Narens, 1990; Raaijmakers & Shiffrin, 1981). Rather, the amount of time subjects were willing to search memory is assumed to be indicated by the latency of the "don't know" responses (cf. Glucksburg & McClosky, 1981; Reder, 1987). Generally speaking, if familiarity is a factor that positively affects the decision to search, the average "don't know" latency for cues that produce a high degree of familiarity should be longer than the average “don’t know” latency for cues that produce a low degree of familiarity. There are, however, several specific hypotheses to consider concerning the effect of familiarity on cued recall performance.

Null hypothesis. The output of the global-matching process has no significant effect on the decision of when to terminate a search and the familiarity manipulation does not produce interference. If the null hypothesis is correct, the familiarity manipulation should not have a significant effect on the mean proportions of a correct response or on the mean response latencies for either correct or “don’t know” responses. For example, Diller et al.’s (2001) REM model does not predict a list-strength effect for cued recall (Shiffrin & Steyvers, 1997; also see Ratcliff, Clark, & Shiffrin, 1990 for the relevant findings concerning list-strength effects for cued recall). Thus, storing relatively strong memory traces does not interfere with retrieval of relatively weak traces.

Effective-search hypothesis. The output of the global-matching process affects the decision of when to terminate a search, additional retrieval attempts increase the chance of success, and either (a) the familiarity manipulation does not produce interference or (b) the additional time spent searching improves recall to a greater extent than interference harms recall. The effective-search hypothesis assumes the additional time spent searching memory will increase the probability of success either because subsequent retrieval attempts with the same set of cues are independent or because cues are changed on subsequent attempts producing additional opportunities to find an effective retrieval cue (cf. Diller et al., 2001). If the effective-search hypothesis is correct, “don’t know” latencies should be longer for cues that produce a relatively high degree of familiarity, and the additional time spent searching memory should produce higher probabilities of correct responses.

There are two possible scenarios involving the latencies of the correct responses that are consistent with the effective-search hypothesis. One is that relatively familiar cues produce longer average latencies for correct responses because some of the extra searches will result in the retrieval of the target. Another result that is consistent with effective-search hypothesis is that cue familiarity may have a countervailing effect on the time course of retrieval by producing some relatively fast correct responses in addition to some relative slow correct responses. That is, the average latency for the earliest correct responses may be shorter for functionally stronger than for functionally weaker cue-target pairs. If so, an increase in correct recall may be observed even though the latencies of correct responses appear to be independent of the familiarity of the cue.

Ineffective-search hypothesis. The output of the matching process affects the decision of when to terminate a search, additional retrieval attempts do not increase the chance of success, and the familiarity manipulation does not produce interference (see above). If the ineffective-search hypothesis is correct, "don't know" latencies should be longer for cues that evoke a relatively high degree of familiarity. In addition, the longer time spent searching memory should have no significant effect on either the probabilities or latencies of correct responses because the extra searches are being carried out with ineffective retrieval cues. For example, access to memory is direct in many composite memory models (e.g., TODAM2, Murdock, 1993; the Matrix Model, Humphries et al., 1989). For this reason, repeatedly probing with the same retrieval cue would not increase the probability of correct recall because the state of memory does not change. If, however, subjects vary the contents of the retrieval cue from one probe to the next, then additional probes may produce an increase in the likelihood of successful retrieval.

Even in a separate-trace global-memory model like SAM or REM, where multiple searches are carried out and different traces may be retrieved due to the stochastic nature of retrieval, additional searches may not necessarily produce a large increase in the probability of correct recall if subjects do not change retrieval cues one from probe to the next. Why might subjects be reluctant to change retrieval cues? In cued recall, the task is to remember the word that was paired with the experimenter-provided cue at study. One variant of the ineffective-search hypothesis assumes that additional memory probes use the same ineffective retrieval cues as earlier probes and that probing memory with the same ineffective retrieval cue produces the same result (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980). It would make little sense from the subject's point of view to abandon the experimenter-provided retrieval cue given the nature of the task.

Interference hypotheses. The familiarity manipulations may produce interference that makes it more difficult to retrieve the target item from memory. Interference is often thought of as a form of response competition that occurs when two or more possible responses are associated with, and produced by, the information in the retrieval cue (see M. C. Anderson & Neely, 1998 for a review). On this basis, interference is expected to produce longer latencies for correct responses because resolving the competition between responses takes time (cf. Anderson ,1981; Goebel and Lewandowsky, 1991), and lower proportions of correct responses because sometimes the incorrect item that is producing the interference will be chosen. However, interference will not affect the latencies of don't know responses.

Experiment 1

An extra-list direct-priming procedure was used to manipulate the familiarity of the cues (Metcalfe et al., 1993). Subjects carried out a series of word-fragment completion trials prior to the presentation of the paired-associate study list. Half of the words designated to be cues at test appeared during the word fragment completion trials (*primed cues*), and the remaining cues only appeared on the study list (*unprimed cues*). If familiarity is a factor influencing the length of search and to the extent that the episodic traces stored during the priming phase take part in the global-matching process, then the latencies of don't know responses to the primed cues will be longer than those in response to the unprimed cues.

Method

Subjects, Design, & Materials. Forty-six introductory psychology students participated in exchange for course credit. A single within-subjects factor, primed versus unprimed cue, was varied. Eighty words were randomly drawn for each subject from a pool of 100 words used for word-fragment completion tasks by Rajaram and Roediger (1993). Forty paired associates were formed for each subject by randomly pairing two words, and one of the words from each pair was randomly selected to be a cue at test.

For each subject, 20 paired associates were randomly assigned to the primed condition, and the remaining 20 pairs were assigned to the unprimed condition. Priming was operationally defined as the presentation of cues prior to study during word-fragment completion trials. The 20 words serving as cues in the primed condition were decomposed into word fragments by removing one letter such that each fragment could be completed to form exactly one word.

The dependent variables of interest were the latencies and probabilities of correct and don't know responses. Latencies of correct responses were measured from the time the cue appeared on the monitor to the time the subject entered the first letter of a response. Don't know latencies were measured from the time the cue appeared on the monitor to the time the subject pressed a key signaling he or she did not remember the target item. With a single exception, the frequencies of incorrect responses were too low to enable meaningful data analyses. Therefore, with the one exception, these data will not be discussed further.

Procedure. The experiment was conducted on personal computers in individual subject booths. Subjects were first given standard instructions about the cued-recall phase of the experiment and were told that they had as long as they wanted to try to remember the target response. They were also told that if they could not remember the word paired with the cue, they could end the current trial at anytime and move on to the next trial by entering a 'don't know' response.

After receiving instructions for the cued recall portion of the experiment, subjects were given instructions for the word fragment completion task. They were told that the purpose of the word fragment completion task was to become familiar with entering responses using the computer keyboard. On each priming trial, a word fragment was displayed in the center of the computer monitor. After the letter that correctly completed the word fragment was entered by the subject, the next priming trial began.

After finishing the word fragment completion trials, subjects were reminded of the cued recall instructions. During the learning phase of the experiment paired associates were presented side-by-side in the center of a computer monitor for 5 s. Upon completion of the study phase, subjects performed a distractor task lasting at least 30 seconds. The distractor task consisted of adding 10 random digits that were presented one at a time at a rate of one every 3 seconds. Cued recall testing followed the distractor task.

On each cued-recall trial, one word from a studied pair was displayed in the center of the monitor. Below the cue, a prompt was displayed where subjects would type in their response to the cue. When subjects thought they knew the word that had been paired with the cue they typed the word on the computer keyboard and pressed "Enter". When subjects thought they did not know the answer, they pressed the question mark key on the keyboard. As soon as either response was made by the subject, the next test trial began. The same procedure was used in the remaining three experiments.

Results and Discussion

The standard of significance is .05 and the statistical analyses of the latencies were performed on the log transformed latencies of the correct and "don't know" responses in order to control for outliers (Ratcliff, 1993). It was not possible to guarantee that each subject would produce every possible type of response in every condition of the experiment; thus, the degrees of freedom that are reported may vary from condition to condition.

The mean proportions and latencies of the various responses are reported in Table 1. The don't know latencies for primed cues were significantly greater than for unprimed cues [$t(44) = 2.16$]. Priming did not significantly affect the proportion of DK responses [$t(45) = .65$]. Priming the cue did not have a statistically significant effect on the proportion of correct responses [$t(45) = .20$] or on their latencies [$t(43) = .11$].

Longer don't know latencies for primed cues suggest that the familiarity produced by the retrieval cue affects the search permission control process. The failure to observe statistically reliable effects of priming on either the proportion or latency of correct responses indicates that interference did not differentially affect the priming conditions and is inconsistent with the effective-search hypothesis. The pattern of data is consistent with the ineffective-search hypothesis. Subjects conducted longer searches in response to the relatively familiar cues, but the extra searches did not produce successful recall.

Experiment 2

The semantic similarity of retrieval cues is used in Experiment 2 to manipulate familiarity. For some cue-target pairs (A-C), a related cue-target pair was studied (A'-D). I refer to these as *similar cues*. The cues of the remaining cue-target pairs were chosen randomly and hence they are only incidentally similar to the rest of words comprising the study list. I refer to these as *dissimilar or randomly similar cues.* Assume that similar cues have more semantic features in common than non-similar cues (Estes, 1994; Hintzman, 1987). According to global-matching theories of recognition, the level of familiarity produced by matching a retrieval cue against the contents of memory is a positive function of the similarity between the retrieval cue and the memory set (Clark & Gronlund, 1996). Dissimilar cues will only tend to match their own trace stored during study. However, similar cues will not only match their own trace, they will also partially match the memory trace corresponding to the study trial with the semantically similar cue. Thus, global-matching models predict that the similar cues will elicit higher levels of familiarity than non-similar cues (Hintzman, et. al., 1994). If familiarity positively affects the length of search, then the don't know latencies for similar cues will be longer than for dissimilar cues.

Method

Forty-three students from introductory psychology courses participated in the experiment in exchange for course credit. A single-factor (semantically similar vs. non-similar cues) within-subjects design was used. Semantic similarity was operationally defined as two exemplars from the same semantic category according to the Battig & Montigue (1968) norms. Sixty paired-associates were randomly formed for each subject. Half of the cues were semantically similar to other cues and half were not. For each subject, 60 target words were randomly assigned to the 60 cues.

Results and Discussion

Four subjects' data were not included in the statistical analysis because of failure to understand the instructions or computer malfunction. The mean proportion and latencies of the different responses are presented in Table 2. Subjects searched longer in response to similar cues than to non-similar cues [$t(38) = 2.67$]. In addition, subjects made significantly fewer don't know responses to similar cues [$t(38) = 2.56$]. The similarity of the cue did not significantly affect the proportions [$t(38) = .70$] or the latencies of correct responses [$t(36) = .98$]. Higher-levels of familiarity were associated with longer searches, and the additional searches did not produce successful retrievals. In fact, cue similarity increased the number of incorrect responses at the expense (i.e., commission errors) of the don't know responses but had no effect on the correct responses. Thus, the additional time spent searching did not improve the accuracy of cued recall; in fact, it was correlated with a lower level of accuracy.

Experiment 3

In this experiment, the familiarity produced by the retrieval cue is manipulated by controlling the amount of time the cue is available for study during the learning phase of the experiment. This is accomplished using an offset-study design (cf. Benjamin, 2005); conditions in which the cue and target appear together for *t* seconds are compared with conditions in which a *t*-second pairing of the cue and target is preceded by an *s*-second presentation of the cue alone. Increasing the amount of time that a cue is studied should increase it familiarity. The design is shown in Figure 1.

For the "short pairs", the cue and target appear simultaneously and remain on screen together for 2.5 s. For the "long pairs", the cue and target also appear simultaneously and remain on screen together for 7.5 s. For the "offset pairs", the cue appears on the screen alone for 5 s. after which it is joined by the target, and the pair remains onscreen together for an additional 2.5 s. Thus, the offset cues are presented for the same amount of time as the long cues, but the offset cues and targets are presented as a pair for the same amount of time as the short pairs.

If cues that evoke higher levels of familiarity produce longer search times, don't know should be longer for offset and long pairs than for short pairs because the offset and long cues should be more strongly encoded. The effective search hypothesis predicts that the additional time searching will increase the proportion of correct responses in these conditions. The ineffective search hypothesis predicts that the additional time spent searching will not increase the proportion of correct responses. The interference hypothesis predicts that the proportion of correct responses will be greater in the short than in the offset and long conditions.

Method

Subjects, Design and Materials. Sixty volunteers from introductory psychology courses participated in exchange for course credit. For each subject, 90 nouns with normative frequencies between 20 and 50 per million were randomly selected from the Kucera and Francis (1967) pool of words used in Experiment 1 and formed into 45 pairs. Pair type was the single within-subjects factor manipulated at three levels: short, long, and offset. For each subject, 15 pairs were randomly selected to serve in each condition, and one word from each pair was randomly selected for each subject to serve as the cue.

Cues were presented simultaneously with the target in both the short- and long-study conditions. Short pairs were studied for 2.5 s., and long pairs were studied for 7.5 s. Offset cues were presented 5.0 s. prior to the presentation of the target, after which the cue and the target were studied for 2.5 s. together. Study order was completely randomized for each subject in order to control for lag. The dependent variables of interest were the latencies and probabilities of correct and don't know responses.

Results and Discussion

The mean latencies and response probabilities are presented in Table 3. The pair-type manipulation had a significant effect on both the latencies [$F(2,114) = 3.70$] and the proportion [$F(2,118) = 7.04$] of don't know responses. Subjects searched longer with offset [$t(57) = 2.41$] and long cues [$t(57) = 2.60$] than with short cues, but the don't know latencies for the offset and long cues did not differ significantly [$t(57) = .29$]. Thus, subjects searched longer to relatively familiar cues.

Subjects made significantly fewer don't know responses in the offset [$t(59) = 2.19$] and long conditions [$t(59) = 3.68$] than in the short condition. The proportions of don't know responses for the offset and long cues did not differ significantly [$t(57) = .33$]. The difference in proportions of don't know responses is complemented by a difference in the proportion of correct responses [$F(2,118) = 4.10$] but not on their latencies [$F(2,114) = 1.49$]. The proportion of correct responses for short pairs was significantly less than for long [$t(59) = 2.70$] and offset pairs [$t(59) = 2.20$], and the latter two conditions did not differ significantly [$t(59) = .70$]. The longer subjects searched memory the greater the proportion of correct responses and the lower the proportion of don't know responses.

The finding that "don't know" latencies for long and offset pairs were greater than for short pairs provides evidence that the search permission control process is positively affected by the familiarity of the retrieval cue. These longer latencies to respond don't know were also associated with increase proportions of correct responses, which suggests that the willingness to spend additional time searching was somewhat effective. The fact the latencies of correct responses didn't differ significantly suggests that increasing the strength with which the cue is encoded decreases the amount of time it takes access at least some traces in memory, offsetting the increased amount of time associated with retrieving other traces from memory.

Experiment 4

In the prior experiments, the familiarity manipulation produced longer memory searches. Experiment 4 examines the question, in an *a priori* manner, of whether the use of familiarity to control search time can be strategically overridden in circumstances in which the subject has reason to believe that familiarity may not be a reliable indicator of memorability. It is identical to Experiment 3 with the exception that the long pairs are eliminated in Experiment 4 leaving only the short and offset pairs.

In Experiment 3, the link between increases in familiarity and study time was salient, but the presence of the long pairs gave subjects reason to believe that familiarity was a reliable indicator of target memorability. Eliminating the long pairs may lead subjects to disregard familiarity as an indicator of memorability because subjects note the amount of time studying the cue is not correlated with the amount of time studying the pair. That is in Experiment 4, the reason why some cues produce higher levels of familiarity than others is salient, but there is also reason to believe that familiarity is not a reliable indicator of the memorability of the target. On these assumptions, removing the long cues in Experiment 4 should result in equivalent don’t know latencies for short and offset pairs. As a result, the proportion correct for the short and offset pairs should also be equivalent.

Method

Subjects, Design and Materials. Thirty-four volunteers from introductory psychology courses participated in exchange for course credit. A single within-subjects factor (short pairs vs. offset pairs) was manipulated in the paired-associate cued recall procedure used in the previous experiments. For each subject, 80 words were randomly drawn from the same pool of words used in Experiment 3 and randomly formed into 40 paired associates for each subject. One of the items from each pair was randomly selected to be a cue at test and the other member of the pair served as the target for the cue. Pairs were randomly divided between the short and offset conditions for each subject. Each short pair of words was studied together for 2.5 s. The offset cues appeared on the computer screen 5 s. prior to the presentation of the target, after which the cue and the target were studied together for 2.5 s.

Results and Discussion

The mean latencies and response probabilities are presented in Table 4. The results are easy to describe: The amount of time the cue was studied did not have a significant effect on any of the dependent measures. The importance of these null results can best be understood in comparison with the results of Experiment 3. The sole difference between Experiments 3 and 4 is the presence of long pairs during the learning phase of Experiment 3, and the absence of these long pairs had two important consequences. The familiarity of the cue did not affect how long subjects were willing to search memory and hence the proportion of correct responses was same for the short and the offset conditions. Apparently, subjects judged that the additional time spent studying the cues in the offset condition relative to the short condition would help them remember the targets, and hence length of search was based on something other than cue familiarity.

I hypothesized that familiarity would be overridden in Experiment 4 for two reasons. First, the source of the familiarity was salient because they knew they had studied the cue by itself during the time it appeared by itself in the offset condition. Second, subjects believed the familiarity was not a good indicator of memorability because the time spent studying the pairs together was the same regardless of how long they studied the cue. One might have expected that improving the encoding of the cues by increasing the amount of time that they were studied would have improved memory in the offset condition regardless of whether subjects were willing to search longer. However, during the time when the cue was presented by itself in the offset condition it might have been not encoded in a manner that strengthened the cue-target association. For instance, the representations of the cue and the cue-target association may have been stored in separate traces (Murdock, 1993), and without additional search time access to the associative trace was not improved.

General Discussion

Other Factors that Might Influence Length of Search

As a package, the results of these experiments suggest that cue familiarity can affect but does not always affect the amount of time one is willing to search memory. When the familiarity of the cue is thought to be correlated with the memorability of the target, relatively familiar cues can produce longer average length of searches and better recall performance. On the other hand, Experiment 2 showed that even when the additional time spent searching produced lower accuracy due to interference, cue familiarity positively affected the length of search. Lastly, when the familiarity of the cue is not thought to be correlated with the memorability of the target, it appears to play little or no role in the determining the length of search.

The final conclusion begs the question: When cue familiarity is not affecting length of search, what is affecting the length of search? It is, of course, quite possible that feeling-of-knowing judgments are at times influenced by factors other than cue familiarity. In fact, a large number of variables have been posited to possibly affect feeling-of-knowing judgments (Nelson et al., 1984).

Koriat (1993) made the general distinction between information provided by an *internal monitor* and *trace accessibility*. The internal monitor is assumed to provide information about the presence versus the absence of an item in memory based on processes that are independent of those used to access memory when performing a recall task, whereas information produced by structural retrieval processes provides clues to the subject as to how accessible an item is. Without further specification of the nature of the internal monitor, this assumption concerning the basis of feeling-of-knowing judgments is rather unsatisfactory on a meta-theoretical basis, and it has been said to be rejected on empirical grounds (Koriat, 1993). Indeed, Koriat preferred the hypothesis that the by-products of unsuccessful retrieval attempts influence feeling-of-knowing judgments. Namely the amount and intensity of the information retrieved from memory are the basis for feeling-of-knowing judgments, and these constructs map nicely onto the global-memory framework that assumes that retrieval processes produce information about specific items in memory and global-matching processes produce information about an item's familiarity (cf. Hintzman, 1987).

This particular trace accessibility hypothesis comes up short, however, when applied to the present results. First, it is unclear why the extra-list cue-priming manipulation used in Experiment 1 would enhance the amount of target information retrieved. Second, the results of Experiment 3 might be explained by assuming that the intensity of the information retrieved from memory only corresponded to that information associated with the cue (i.e., cue familiarity) and that only the intensity of the information retrieved from memory was used to guide length of search. If one assumes that accessibility of the target trace is what governs length of search, then one would have expected longer average length of searches in the long-pair condition relative to the offset-pair condition since the targets were studied much longer in the long-pair condition and hence more information about them should have been accessible. Moreover, this cue-familiarity version of the trace accessibility hypothesis cannot explain why eliminating the long-pairs from the study list, as was done in Experiment 4, produces similar search durations for relatively familiar and unfamiliar cues.

It appears that length of search, at least at times, can be influenced by factors that have little to do with how accessible items are. For instance, given the results of Experiment 3, we would have expected for recall to better in the offset condition of Experiment 4 if subjects had been willing to search longer. Nelson et al. (1984) discussed several other factors that could affect feeling-of-knowing judgments and perhaps length of search. They made a distinction between *trace-access* mechanisms and *inferential* mechanisms. According to Nelson et al., "trace-access mechanisms share the characteristic that the person is presumed to have access to nonrecalled item during feeling of knowing judgments.", whereas for inferential mechanisms "the feeling of knowing does not monitor the nonrecalled target item". Nelson et al. assigned a large number of possible mechanisms to one or the other classes that could give rise to a feeling-of-knowing judgment. For instance, the retrieval of different types of partial information was classified as a trace access mechanism, whereas cue familiarity was classified as an inferential mechanism.

Several, other trace access and inferential mechanisms were discussed by Nelson et al. (1984), but given the current state of the science of structural memory theory some of the distinctions between trace access and inferential mechanisms are a bit blurry. For instance, producing cue familiarity involves access to the contents of trace representing the cue, even if those contents are not available to the subject. More generally, one might define a trace access mechanism as one that provides information about a particular aspect of an item in memory, whereas an inferential mechanism provides information that is not specific to any particular item. The later type of information could be used to affect the length of search for a particular cue based on what is known or believed about the typical item or class of items. Such a conceptualization of trace access is more consistent with Koriat's (1993) model while preserving Nelson et al.'s (1984) notion of the possibility that other factors can affect feelings-of-knowing or length of search.

In the present case, for instance, it seems plausible that subjects learned something about the nature of the study list as a whole in addition to the individual word pairs that comprised it. That is, in Experiment 3 subjects might have noticed that cue strength was positively (if not perfectly) correlated with target strength, whereas in

Experiment 4 they were independent of each other. When combined with a heuristic that states that the familiarity of the cue is a valid predictor of successful recall only when it is positively correlated with strength with which the target is encoded, subjects may choose is utilize cue familiarity as a determinant of length of search.

On the Accuracy of Feeling of Knowing Judgments

In addition to the factors that affect feeling-of-knowing judgments and length of search, a critical question has to do with why feeling-of-knowing judgments are only moderately predictive of subsequent criterial testing performance (cf. Nelson & Narens, 1990). Koriat (1993) proposed that trace access mechanisms might provide information that leads either to correct or incorrect feeling-of-knowing judgments. Because subjects have no direct way of assessing the validity of the information retrieved from memory, feeling-of-knowing judgments can be mislead. On the other hand, memory strength or familiarity has no direct influence on feeling-knowing judgments, but is simply assumed to be correlated with the amount of partial information that is retrieved about the target such that increases in memory strength produce more correct partial information and less incorrect partial information, leading to a positive correlation between feeling-of-knowing judgments and recognition performance.

The assumption that memory strength and the retrieval of partial information are correlated is called into question by factors that have opposite effects on recognition and recall, such as word frequency (Gillund & Shiffrin, 1984). In addition, two findings from Experiments 1 and 2 call into question the assumption that familiarity does not have a direct effect on feeling-of-knowing judgments. In Experiment 1, some of the cues used in cued recall phase were presented prior to the study list as a part of a word-fragment completion task. Later, when cued recall was tested, subjects were willing to search longer when cued with a previously primed word. In Experiment 2, the study list consisted of some cues that were only randomly similar to the other cues on the study list, whereas the remaining cues were semantic associates of another cue on the study list. Because familiarity is assumed to be a positive function of the similarity between a retrieval cue and the contents of memory (i.e., the target trace and the traces of other studied items) semantically similar cues should have seemed more familiar at test than randomly similar cues. The finding that semantically similar cues produced longer average lengths of search confirmed these assumptions. While these findings are consistent with a cue familiarity hypothesis, it is difficult within a global memory framework to explain why these operations would have led to increases in the amount of partial target information retrieved.

Here I propose that the relatively moderate correlations between feeling-of-knowing judgments and recognition accuracy might be the result of at least three factors. First, methodological factors can negatively affect feeling-of-knowing judgments. Typically, feeling of knowing judgments are only obtained after unsuccessful attempts to recall. However, subjects presumably have access to the types of information used to make feeling-of-knowing judgments even when recall was successful. In these cases, one would expect that the feeling-of-knowing judgments are much better predictors of recognition performance.

Second, feeling-of-knowing judgments based on inferential mechanisms might be mislead and/or the heuristic used might not a valid. For instance, one might expect that feeling-of-knowing judgments made in the offset condition in Experiment 2 would be less predictive of recognition than those made in the same condition of Experiment 1. Confirmation of this rather speculative hypothesis must wait for further experimentation.

Lastly, the accuracy of feelings-of-knowing judgments might be negatively influenced by cue familiarity. As mentioned in the introduction to this chapter, structural theories of memory typically assume that a global-matching process is responsible for producing a sense of familiarity associated with the nominal cue. The global-matching assumption assumes that the retrieval cue is compared to many traces in memory in addition to the target trace. This produces a somewhat noisy result as the spurious matches or mismatches influence the familiarity that results from memory access. To the extent that spurious matches provide misleading levels of cue familiarity, one expects that feeling-of-knowing judgments to be inaccurate predictors of subsequent recognition performance.

Conclusions

This endeavor is relatively unusual because it acknowledges the contributions of both structural and metamemory research by combining them in a single project that investigates the controlled use of human memory. There remain many issues to investigate concerning the interaction of structural and metamemory processes, and I hope that this research provides a reasonable example of they might be addressed.

The present experiments were jointly motivated by common assumptions made by structural memory and metamemory theories. I was particularly intrigued by the possibility of gathering relevant observations that could help extend extant memory models to the temporal dynamics associated with retrieval, an issue that is usually ignored for sake of simplicity. I was also intrigued by the possibility of constraining several hypotheses concerning length of search made in the metamemory literature by several well-supported assumptions made by structural memory models. Based on these assumptions, the present results support the notion that cue familiarity can affect how long one is willing to search memory but only when cue familiarity is not attributed to spurious factors. In addition, the length of search appears to be only incidentally related to its effectiveness.

References

Anderson, J. R. (1981). Interference: The relationship between response latency and response accuracy. Journal of Experimental Psychology: Human Learning and Memory, 7(5), 326-343.

Atkinson, R. C. & Juola, J.F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (ed.), Attention and Performance (Vol. 4). New York, Academic Press.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed systems and its control processes. In W. E. Spence & J. T. Spence (Eds.), The psychology of learning and motivation (2), New York: Academic.

Clark, S. E. (1998). Recalling to recognize and recognizing recall. In Izawa, C. (Ed.), On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model. Hillsdale, NJ.: Earlbaum.

Clark, S. E. & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin & Review, 3(1), 37-60.

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, & Cognition, 27*, 414-435.

Estes, W. K. (1994). Classification and Cognition. New York: Oxford.

Gardiner, J. M. & Java, R. I. (1990). Recollective experience in word and non-word recognition. Memory & Cognition, 18, 23-30.

Gillund, G. & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Glucksburg, S., & McCloskey, M. (1981). Decision about ignorance: Knowing that you don't know. Journal of Experimental Psychology: Human Learning and Memory, 7, 311-325.

Goebel, R. P. & Lewandowsky, S. (1991). Retrieval measures in distributed memory models. In Hockley, E., & Lewandowsky, S. (Eds.), Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock. Hillsdale, NJ: Earlbaum.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1994). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. Memory & Cognition, 26(3), 449-462.

Hintzman, D. L. (1987). Recognition and recall in MINERVA2: Analysis of the "recognition failure" paradigm. In P. Morris (Ed.), Modeling cognition: Proceedings of the international workshop on modeling cognition. London: Wiley.

Humphreys, M. S., M. Bain, J. D., & Pike, R. (1989). Different way to cue a coherent memory system: A theory of episodic, semantic, and procedural tasks. Psychological Review, 96, 208-233.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. Psychological Review, 100, 609-639.

Kucera, H., & Francis, W. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on judgments of frequency and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 319-331.

Malmberg, K. J. & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognitio*n, 31(2) 322-336.

Malmberg, K. J. & Xu, J. (in press). On flexibility and on the fallibility of associative memory, *Memory & Cognition*.

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R.M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by Midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 540-549.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87(3), 252-271.

Mandler, G. (1991). Your face looks familiar but I can't remember your name: A review of dual-process theory. In Hockley, W. E., & Lewandowsky, S. (Eds.), Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock. Hillsdale, NJ: Earlbaum.

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holograph associative recall model: Implication for Korsakoff Amnesia. Psychological Review, 100(1), 3-22.

Metcalfe, J. M., Schwartz, B. L., & Joaquim (1993). The cue-familiarity heuristic in metacognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(4), 861-861.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. Psychological Review, 100(2), 183-203.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. Journal of Experimental Psychology: General, 113, 282-300.

Nelson, T. O. & Narens, L. (1990). Metamemory: A Theoretical framework and new findings. In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory. New York: Academic Press.

Raaijmakers, J. G. W. & Shiffrin, R. M. (1981). Search of associative memory. Psychological Review, 88(2), 93-134.

Raaijmakers, J. G. W. & Shiffrin, R. M. (1992). Models for recall and recognition. The Annual Review of Psychology, 43, 205-234.

Rajaram, S. & Roediger, H.L. (1993). Direct comparison of four implicit memory tests. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(4), 765-76.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological Bulletin, 114, 510-532.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect I: Data and discussion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 162-178.

Reder, L. M. (1987). Strategy Selection in question answering. Cognitive Psychology, 12, 90-138.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294-320.

Schwartz, B. L. & Metcalfe, J. M.(1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18**,** 1074-1083.

Shiffrin, R. M., Murnane, K., Gronlund, S., & Roth, M. (1989). On units of storage and retrieval. In Izawa, C. (Ed.), Current issues in cognitive processes: The Tulane Floweree Symposium on cognition. Hilldale, NJ.: Earlbaum.

Shiffrin, R.M. & Steyvers (1997). A model for recognition memory: REM: retrieving effectively from memory. Psychonomic Bulletin & Review, 4(2), 145-166.

Snodgrass, J. G. (1987). Discussion of chapter by Murdock. In Gorfein, D. S. and Hoffman, R. R. (Eds.), Memory and learning: The Ebbinghaus centennial conference, Hillsdale: Earlbaum.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Figure 1. *Pair Types that were Used in the Designs of Experiments 1 Versus 2*

Experiment 1:

	Cue Only	Pair
Offset Pairs	5.0 s.	2.5 s.
Short Pairs		2.5 s.
Long Pairs		7.5 s.

Experiment 2

	Cue Only	Pair
Offset Pairs	5.0 s.	2.5 s.
Short Pairs		2.5 s.

Table 1: *Mean Proportions and Latencies of Correct Responses and Don't Know Responses for Experiment 1*

Priming Condition	Response Type			
	Correct Responses		Don't Knows	
	Proportion	Latency (s.)	Proportion	Latency (s.)
Primed Cue	.42	3.8	.45	7.1
Unprimed Cue	.42	3.5	.47	6.1

Note: The proportions of correct and don't know responses do not sum to 1.0 due to the fact that commission errors were sometimes made.

Table 2: *Mean Proportions and Latencies of Correct Responses and Don't Know Responses for Experiment 2*

Priming Condition	Response Type			
	Correct Responses		Don't Knows	
	Proportion	Latency (s.)	Proportion	Latency (s.)
Similar Cue	.21	3.5	.62	5.5
Dissimilar Cue	.20	3.7	.68	4.8

Note: The proportions of correct and don't know responses do not sum to 1.0 due to the fact that commission errors were sometimes made.

Table 3: *Mean Proportions and Latencies for Correct and Don't Know Responses for Experiment 3.*

	Response Type			
	Correct Responses		Don't Know Responses	
Pair Type	Proportion	Latency (s.)	Proportion	Latency (s.)
Short	.32	3.3	.60	4.2
Offset	.37	3.1	.53	4.9
Long	.39	3.0	.53	5.0

Note. The proportions of correct and don't know responses do not sum to 1.0 due to the fact that commission errors were sometimes made.

Table 4: *Mean Proportions and Latencies for Correct and Don't Know Responses for Experiment 4*

	Response Type			
	Correct Responses		Don't Know Responses	
Pair Type	Proportion	Latency (s.)	Proportion	Latency (s.)
Short	.33	3.1	.56	4.2
Offset	.35	3.3	.54	4.2

Note. The proportions of correct and don't know responses do not sum to 1.0 due to the fact that commission errors were sometimes made.

[1] Although the use of familiarity in recall has not been widely examined, it has not been ignored. Some composite storage memory models like CHARM (Metcalfe-Eich, 1982) and TODAM (Murdock, 1982) posit that a matching process is involved in a post-retrieval deblurring process that is used to eliminate noise from the retrieved content in cued recall (see Goebel & Lewandowsky, 1991; and Snodgrass, 1987 for critiques). The noisy output is matched against a lexicon of possible responses and the highest match is chosen as the response. In SAM and REM (Diller, Nobel, and Shiffrin, 2001; Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), sampling probability for recall is based on the similarity of the retrieval cues and traces relative the normalized to global-match strength.