

2006

How cue-dependent is memory?: Internal reinstatement and cueing effects in recognition and source memory

Jeffrey Joseph Starns

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_dissertations



Part of the [Psychology Commons](#)

Recommended Citation

Starns, Jeffrey Joseph, "How cue-dependent is memory?: Internal reinstatement and cueing effects in recognition and source memory" (2006). *LSU Doctoral Dissertations*. 860.

https://repository.lsu.edu/gradschool_dissertations/860

This Dissertation is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.

HOW CUE-DEPENDENT IS MEMORY?
INTERNAL REINSTATEMENT AND CUEING EFFECTS IN RECOGNITION AND
SOURCE MEMORY

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by

Jeffrey Joseph Starns

B.A., Southeastern Louisiana University, 2001

M.A., Louisiana State University, 2004

August 2006

TABLE OF CONTENTS

ABSTRACT	iv
INTRODUCTION	1
Context Effects in Recognition and Source Memory	1
Explanations for Null Context Effects	4
A Comprehensive Theory for Context Effects	8
EXPERIMENT 1	12
Simulation Model for Source Memory in Experiment 1	13
Experiment 1 Simulation Results	21
An Alternative Internal Reinstatement Strategy	26
Experiment 1 Method	26
Participants	26
Power Analysis	27
Design	27
Materials	28
Procedure	30
Experiment 1 Results and Discussion	31
EXPERIMENT 2	36
Simulation Model for Recognition Memory in Experiment 2	36
Experiment 2 Simulation Results	37
Alternative Internal Reinstatement Strategy	39
Experiment 2 Method	40
Participants	40
Design	40
Materials	40
Procedure	40
Experiment 2 Results and Discussion	41
EXPERIMENT 3	45
Experiment 3 Method	46
Participants	46
Design	46
Materials	46
Procedure	47
Experiment 3 Results and Discussion	48

EXPERIMENT 4	51
Experiment 4 Method	52
Participants	52
Design	52
Materials	52
Procedure	52
Experiment 4 Results and Discussion	53
GENERAL DISCUSSION	56
Consistency with the Results of Existing Studies	59
Global Matching Models and Episodic Memory	63
Conclusion	65
REFERENCES	67
VITA	71

ABSTRACT

This study explored the role of internal context reinstatement in masking the effects of external context cues on recognition and source memory. Participants studied words paired with pictures of male and female faces. Following the study phase, participants completed either a source test in which they decided whether each test word was studied with a male or female face (Experiments 1, 3, and 4) or a recognition test in which they decided whether each test word appeared in the study phase (Experiment 2). On selected trials, a studied face was reinstated at test to serve as a cue for the memory decision. In each experiment, this cueing manipulation was factorially crossed with a manipulation designed to impair participants' ability to internally construct appropriate face cues when no face was externally reinstated at test. In Experiments 1 and 2, separate groups of participants studied either sets of similar faces (high-overlap condition) or sets of distinct faces (low-overlap condition). MINERVA 2 simulation models showed that internal reinstatement was less effective in the low-overlap condition; consequently, external face cues significantly improved performance in the low-overlap condition but not the high-overlap condition. In contrast, the empirical results showed no cueing effects in either the high- or low-overlap conditions. In Experiments 3 and 4, separate groups of participants studied either a single male and female face or multiple male and female faces. Results showed that external cues improved memory performance for the multiple-face participants, but did not influence performance for participants who studied a single male and female face. This pattern is consistent with the hypothesis that presenting multiple faces disrupted participants' ability to internally reinstate appropriate face cues, thus making performance more dependent on external cues.

INTRODUCTION

An emphasis on the importance of retrieval cues permeates current theories of long-term memory. It is generally accepted among memory theorists that long-term forgetting primarily reflects a lack of appropriate retrieval cues, whereas the loss or decay of stored information is less important (Tulving, 1974). Moreover, general theoretical principles [e.g., encoding specificity (Tulving & Thomson, 1973)] and specific mathematical models of memory [e.g., MINERVA 2 (Hintzman, 1988)] postulate that memory performance depends on the match between processing at encoding and retrieval, with better memory predicted when there is greater overlap in processing. From this perspective, retrieval cues dramatically affect memory by promoting or discouraging the reinstatement of encoding processes at retrieval.

Context reinstatement is frequently used to explore the effects of retrieval cues (see Smith & Vela, 2001, for a review). If contextual details are typically processed and incorporated into long-term memories at encoding (Johnson, Hastroudi, & Lindsay, 1993), then reinstating aspects of the encoding context at retrieval should promote the overlap of encoding and retrieval processes, thereby improving memory performance. Studies using recall tests convincingly demonstrate that context reinstatement improves memory performance (Smith & Vela, 2001). In a classic example, participants recall more words when they are tested in the same room in which they learned the words than when they are tested in a different room (e.g., Smith, 1979).

Context Effects in Recognition and Source Memory

In contrast to the robust context effects produced in recall, null effects of context reinstatement are regularly observed in studies using recognition or source-monitoring tests (Craik & Kirsner, 1974; Fernandez & Glenberg, 1985; Godden & Baddeley, 1980; Kirsner, 1973; Murnane & Phelps, 1993, 1994, 1995; Smith, Glenberg, & Bjork, 1978; Starns & Hicks,

2004; Starns & Hicks, 2005). On a recognition test, participants are asked to discriminate words they saw in a previous phase of the experiment from words that were not previously presented. On a source test, participants are asked to discriminate words previously presented in one context (e.g., a male voice) from words presented in another context (e.g., a female voice).

The recognition literature provides many examples of null context effects. Smith et al. (1978) and Fernandez and Glenberg (1985) showed no differences in recognition memory for participants tested in the room where they learned a list of words versus a different room. Godden & Baddeley (1980) had scuba divers learn a list of words either on land or under water, and found no differences in recognition memory for participants tested in the learning environment versus participants tested in the alternative environment. In a series of studies, Murnane & Phelps (1993, 1994, 1995) explored the effects of reinstating visual contexts defined in terms of word color, background color, and the position of words on the screen. They consistently found that reinstatement of studied contexts increased positive recognition responses to both studied words (targets) and non-studied words (lures), but did not influence memory performance in terms of the discriminability of targets from lures. Finally, in a study conducted by the current author, participants learned words in either a large or small font size and on either the top or bottom of the computer screen (Starns & Hicks, 2005). Neither font-size reinstatement nor location reinstatement influenced recognition discriminability.

Other studies have shown an effect of context reinstatement on recognition memory, although the effects are generally small. Participants instructed to rehearse words in familiarized male or female voices showed better recognition memory when they heard test words in the rehearsed voice versus a different voice (Geiselman & Bjork, 1980; Geiselman & Glenny, 1977). Smith (1986) showed better recognition memory for words learned and tested in the same versus

a different room when participants learned words as part of a short-term memory task, but no context effects for participants who were aware that a long-term memory test would be administered. Kirsner (1973) and Craik and Kirsner (1974) tested participants' memory in a continuous recognition paradigm, with repeated words either matching or mismatching their prior presentation in terms of font case (Kirsner, 1973) or speaker's voice (Craik & Kirsner, 1974). Both types of reinstatement significantly enhanced participants' ability to identify repeated words, but the cueing effects were quite small. Finally, Murnane et al. (1999) reported a study in which participants studied words along with complex visual scenes, such as a classroom scene with a teacher, a desk, and a blackboard. Reinstating the scenes at test significantly increased the discriminability of targets from lures.

Research on context reinstatement in source monitoring tasks mirrors the equivocal results from the recognition literature. Several studies show no effect of context reinstatement on source memory, even when the reinstated contextual detail forms the basis of the source discrimination. In the continuous recognition tasks employed by Kirsner (1973) and Craik and Kirsner (1974), there were no effects of font case reinstatement on the ability to determine if a word was previously seen in an uppercase or lowercase font (Kirsner, 1973), and no effects of voice reinstatement on accuracy in judging whether a word was previously heard in a male or female voice (Craik & Kirsner, 1974). Work by the current author showed that reinstating the font size of a word did not improve source memory for font size, and reinstating the location of a word did not improve source memory for location (Starns & Hicks, 2005). Another study conducted by the current author showed no effect of voice reinstatement on voice memory, even with encoding tasks that encouraged elaborative encoding of voice details (Starns & Hicks, 2004).

Just as in the recognition literature, there are a few studies demonstrating context reinstatement effects on source monitoring tests. Dodson and Shimamura (2000) presented words in either a male or female voice and then tested the words either in the same voice heard at encoding, in the other studied voice, in a voice that was not heard in the study phase, or with no voice at test. They found that source memory for studied voices was more accurate when words were tested in a voice that matched the words' prior presentation as compared to the other conditions. Palmeri, Goldinger, and Pisoni (1993) similarly showed that voice reinstatement improved participants' voice memory.

Clearly, the existing literature on context reinstatement does not provide consistent evidence that contextual cueing influences performance on recognition or source monitoring tasks, and the equivocal results seem wholly inconsistent with the integral role of retrieval cues in many memory theories. The goal of the present project was to develop a detailed theoretical framework for predicting when context effects should and should not occur, and to establish experimental procedures that promote the consistent observation of context effects. Several theorists have attempted to explain the ineffectiveness of context cues, and the hypotheses of these theorists are reviewed and evaluated in the next section.

Explanations for Null Context Effects

The outshining hypothesis (Smith & Vela, 2001) proposes that context cues will not affect memory performance when powerful alternative cues are available at retrieval. Notably, on recognition and source tests, the actual items presented at encoding are reinstated at test, and the usual advantage of recognition over recall memory demonstrates that this item information provides a powerful retrieval cue (e.g., Koriat & Goldsmith, 1994; Mandler, Pearlstone, & Koopmans, 1969). Thus, when item cues are presented at test, the relative influence of context

cues may become insignificant; that is, the item cues may *outshine* the context cues. Although some empirical evidence can be interpreted in favor of the outshining hypothesis (Smith & Vela, 2001; but see Murnane & Phelps, 1995), this theoretical perspective requires the awkward assumption that context cues are of minor importance on tests in which performance is driven exclusively by the ability to discriminate items encountered in different contexts. Indeed, as will be seen, the theoretical perspective advanced in the current study implies that context information must be utilized at retrieval to discriminate items on recognition and source tests.

Smith and Vela (2001) also suggest that context cues may fail to influence memory performance because context information is not encoded in memory traces. The *overshadowing hypothesis* explains why people may sometimes fail to encode context information. According to this hypothesis, context information is processed under normal conditions; however, contextual processing may be suppressed when ongoing tasks require demanding conceptual processing. In other words, people reallocate the processing resources that are normally devoted to context encoding when they are faced with a particularly difficult task. Suppressing contextual processing at encoding naturally leads to null context effects, because context reinstatement cannot aid in the retrieval of memory traces that lack context information. This hypothesis is a plausible explanation of null context effects in recognition; however, the hypothesis fails to explain the results of studies showing null context effects in source memory. In source memory studies, accurate source discrimination proves that the context details defining source membership are incorporated into memory traces; nevertheless, several studies have found that reinstating these same context features had no impact on memory performance (Craik & Kirsner, 1974; Kirsner, 1973; Starns & Hicks, 2004; Starns & Hicks, 2005).

The ICE model (Murnane et al., 1999) implies that the effectiveness of context reinstatement depends on the manner in which context details are encoded. This model predicts memory performance in terms of three types of information: Item information, Context information, and Ensemble information. Ensemble information results when item and context features are integrated into a unified representation; for example, if the word “dog” is studied in the context of a classroom scene, a participant can create ensemble information by imagining a dog teaching a class (Murnane et al., 1999). Each type of information is retrieved independently in the ICE model, so unintegrated context information cannot serve as a cue for item information. Cueing is observed between context and item information only when the two have been integrated into an ensemble. Murnane et al. (1999) provide empirical evidence for the critical role of ensemble information by showing that contexts that are easy to integrate with item information produce cueing effects, but contexts that are difficult to integrate with item information do not. Although the ICE model provides a plausible explanation of null context effects in recognition, this theory fails to explain null context effects in source memory. In the source memory experiments showing null context effects (Craig & Kirsner, 1974; Kirsner, 1973; Starns & Hicks, 2004; Starns & Hicks, 2005), participants’ ability to remember the specific contexts that were paired with items demonstrates that they integrated context and item information at encoding. Nevertheless, reinstating context cues had no impact on memory performance in these studies.

Theorists have also explained null context effects by claiming that participants internally reinstate context cues when external cues are unavailable at retrieval (Bjork & Richardson-Klavehn, 1989; Murnane & Phelps, 1994; Smith, 1979). External cues may be ineffective because the information in external cues is redundant with information internally reinstated by

participants on uncued trials. I will call this account of null context effects the *cue-discrepancy* explanation, as it predicts that context effects are determined by the discrepancy in information available through external and internal reinstatement. A study by Smith (1979) demonstrates the ability of internal reinstatement to mask external reinstatement effects. In his Experiment 2, Smith had participants recall words either in the same room in which the words were learned or in a different room. The different-context participants were split into three groups: one group was told to try to remember the study room and use it to help them recall words, one group was told to remember the study room and was also given a picture of the study room, and one group received instructions that did not mention the study room (as in the standard paradigm). Results showed that same-context participants remembered more words than different-context participants given standard instructions, but both different-context groups instructed to remember the study room recalled as many words as the same-context group. Thus, internal reinstatement was as effective as external reinstatement and completely masked external cueing effects.

In a third experiment, Smith (1979) attempted to manipulate the effectiveness of internal reinstatement by varying the number of contexts that competed for retrieval with the study room context when internal reinstatement was attempted. On the first day of the experiment, participants performed five tasks, one of which was learning a word list. For one group of participants, the word list was learned in one room and all other tasks were performed in a single different room (2-context group). For another group of participants, the word list was learned in one room and the remaining four tasks were performed in four different rooms (5-context group). On the second day of the experiment, half of the participants in each group performed free recall in the learning room, and half performed free recall in a different room with instructions to try to remember the learning room to help them recall words. Results showed no significant difference

between same- and different-context participants for the 2-context group, but a marked advantage for same- over different-context participants in the 5-context group. With only two study contexts, it was easy for participants to remember the unique details of the encoding room, providing them the opportunity to internally reinstate context information in the different-context condition. Increasing the number of contexts encountered on the first day made it difficult for participants to specifically remember the encoding room; therefore, they could not use internal reinstatement to compensate for the lack of external context cues.

Smith's (1979) results convincingly demonstrate that internal reinstatement can eliminate external cueing effects; however, participants in his experiment had to be instructed to use internal reinstatement, and external reinstatement effects were robust when no special instructions were given. To explain null cueing effects in recognition and source memory, one must assume that participants routinely reinstate context details on these types of test even if they are not instructed to do so. In the next section, I present a formal model that assumes that internal context reinstatement is an intrinsic component of recognition and source memory. I used this model to explore the role of internal reinstatement in producing null context effects.

A Comprehensive Theory for Context Effects

Predictions for source and recognition performance were derived from the simulation model MINERVA 2 (Hintzman, 1988), which is one of a class of theories called global matching models (for a review, see Clark & Gronlund, 1996). The basic tenets of global matching models are that 1) retrieval is accomplished by matching the features of a memory probe to the features of *all* stored memory traces; 2) the probe creates a certain amount of activation determined by the degree of overlap in the features of the test probe and the features in memory traces; 3) this unitary activation value is interpreted as the amount of evidence that the test probe accurately

corresponds to past experience; and 4) participants will claim to remember a test probe if the activation that it inspires passes a criterion value. The global matching perspective is clearly applicable to context reinstatement. First, from this perspective, reinstatement profoundly affects memory performance because retrieval is determined by the match of the memory probe to traces of past events. The cues incorporated into the retrieval probe dramatically impact the probe's similarity to stored information. Second, because retrieval probes are matched in parallel to all of memory, global matching models rely on context information to produce memory for specific events. Without context information, attempts to retrieve information from a specific episode would be thwarted by severe interference from similar events. However, if similar events were experienced in different contexts than the target event, then episodic retrieval could be achieved by incorporating context cues in the retrieval probe (Murnane & Phelps, 1994).

Based on the above discussion, the equivocal context reinstatement results in recognition and source memory seem particularly problematic for global matching models; however, these models can successfully accommodate null context effects based on their description of how participants make recognition and source judgments when no explicit context cues are provided. The instructions of an episodic memory test inform participants that they will be asked to select items that were presented in a particular context (i.e., the general encoding context for a recognition test or a specific context within the study phase for a source test). Participants use the instructions to retrieve/construct a context cue appropriate for the task presented to them (Dennis & Humphreys, 2001). For example, when asked to select items previously spoken in a male voice, participants may construct a context cue including the auditory characteristics of the male voice that they heard in the study phase. On a given test trial, participants form a retrieval probe that combines the context cue with a cue representing the current test item, and the

activation created by this probe is interpreted as the amount of evidence that the item was previously encountered in the target context (Humphreys, Bain, & Pike, 1989). This retrieval process may explain why external context cues sometimes have no effect on recognition or source memory performance: Participants must always use a context cue to perform in these tasks, so the information in external context cues may be completely redundant with the information in internally-reinstated context cues constructed by participants (see Bjork & Richardson-Klavehn, 1989). Thus, the global matching model used in this study is consistent with the cue discrepancy explanation, because the model assumes that external cues only increase performance when they include contextual details that are not internally reinstated.

The goal of the current study was to establish the true effects of context reinstatement on source and recognition memory by influencing the effectiveness of internal reinstatement. In this study, participants learned words that appeared along with either male or female faces. At test, participants were either asked to determine if each word was studied with a male or female face (Experiments 1, 3, and 4) or to determine whether each word was on the study list (Experiment 2). Studied faces were reinstated on some test trials to evaluate cueing effects. In each experiment, a variable designed to influence the effectiveness of internal reinstatement was factorially crossed with the cueing variable. In Experiments 3 and 4, separate groups of participants encountered either a single male and female face or multiple male and female faces in the encoding phase. When participants study multiple faces for each gender, they should not be able to internally reinstate all of the studied faces on each test trial. In contrast, with only one face for each gender, internal reinstatement of the studied faces is a much more manageable task. In Experiments 1 and 2, all participants saw multiple male and female faces, but the faces within each gender had either low or high overlap in their component features. Internal reinstatement

should be more effective when faces have high feature overlap, because participants can use a single face cue that will closely capture the features of each individual face within a gender. In contrast, when faces have low feature overlap, participants will not be able to effectively use a single cue to summarize all of the faces. Based on the cue discrepancy hypothesis, the external reinstatement manipulation should have a larger effect on memory performance when internal reinstatement is less effective; that is, in the low-overlap versus the high-overlap conditions of Experiments 1 and 2 and in the multiple-face versus the single-face conditions of Experiments 3 and 4.

EXPERIMENT 1

In this experiment, participants encoded words paired with four male and four female faces. For separate groups of participants, the faces within each gender had either high or low overlap in features. Feature overlap was manipulated by holding particular facial features constant across all faces within a gender set or allowing the features to vary from face to face. For example, in the high-overlap condition all of the male faces had the same eyes, and in the low-overlap condition each male face had unique eyes. At test, participants were re-presented with the studied words and asked to decide if each was seen with a male or a female face in the study phase. Words on the test appeared either without a face, with the face that was paired with the word at encoding, or with a face of the opposite gender.

The multitude of faces in this experiment should prevent participants from internally reinstating all of the individual studied faces at test (Murnane & Phelps, 1994; Smith, 1979). However, from the global matching perspective, retrieval probes must contain face information to discriminate male and female items. In the next section, I describe a formal simulation model that assumes that participants construct a single face cue that summarizes the critical features of all of the faces in each gender (i.e., an “average” face for the gender). I used the simulation model to explore how manipulating feature overlap impacts the effectiveness of the summary cue (i.e., the effectiveness of internal reinstatement). The simulation model was used only to make qualitative predictions regarding the effects of manipulations, not to quantitatively fit the empirical data from our paradigm (Hintzman, 1988). The model is nothing close to an exact replica of the cognitive processes involved in our experiment; rather, it is a computational system that has the same general properties that the theory under investigation ascribes to the human memory system (namely, that memory decisions are based on the global match of the features in

a test probe to the features stored in memory, that memory probes in episodic tasks must contain a cue representing the target context, and that participants internally reinstate context cues when external cues are unavailable). The model was used to better understand the effects of variables on such a system.

Simulation Model for Source Memory in Experiment 1

Predictions from MINERVA 2 (Hintzman, 1988) were derived via computer simulation using E-Prime programs (Psychological Software Tools, 2002). In MINERVA 2, the information available from an event is represented as a vector of feature values. When an event is experienced, the features of the event are copied into a memory trace with probability L , which is the model's learning parameter. When L is close to 1, nearly all event information is copied into memory. When L is close to 0, very little event information is copied into memory.

Features in a memory trace have a value of 1, -1, or 0, with 0 indicating that no information regarding that feature was stored at encoding. Retrieval is initiated by a probe, which is itself a vector of feature values. The probe creates a certain amount of activation based on its similarity to the traces stored in memory. The similarity of the retrieval probe to each trace is computed by multiplying each feature in the trace by the corresponding feature in the probe, summing all of these products, and dividing by the number of features with a non-zero value in either the probe or the trace. Thus, the equation giving the similarity of a trace to a probe is

$$S = \sum_{j=1}^N P_j T_j / N$$

where P_j is the value of feature j in the probe, T_j is the value of feature j in the trace, and N is the number of features with a value other than zero in the trace and/or the probe. The activation of a trace is computed by cubing its similarity value. The total activation created by a memory probe

is determined by adding the activation values for all traces in memory. This index is known as echo intensity, and is computed by the formula

$$I = \sum_{i=1}^M A_i$$

where A_i is the activation of trace i , and M is the number of traces in the memory set.

Figure 1 provides a basic example of how MINERVA 2 can be used to simulate memory performance. The top of Figure 1 shows two events represented by unique combinations of feature values. The next section shows example memory traces of the two events resulting when the learning parameter (L) is .5. Comparing the trace vectors to the event vectors reveals that some of the non-zero event features are copied into the trace representing the event, while other non-zero event features are set to zero in the trace to reflect unsuccessful encoding. Each non-zero event feature has an independent probability L of being accurately represented in memory (event features with a value of zero are always represented with a zero in memory traces, because information not present in an event cannot be encoded into memory). The final section in Figure 1 shows the retrieval computations involved in deciding if Event A was previously experienced. The features of Event A are reinstated in the retrieval probe, and this probe is matched against all memory traces (only two in this example) to derive an echo intensity value. The activation value for each memory trace appears immediately below the trace. The activation values are determined by the overlap in features between the trace and the probe, and the values are computed by cubing the result of the similarity formula. The memory trace of Event A has a high degree of feature overlap with the probe; consequently, this trace has a relatively high activation value (.187). The memory trace of Event B has only random overlap with the probe; consequently, this trace has a near-zero activation value (.001). The activation of each trace is

Event Features

$$\text{Event A: } \begin{pmatrix} 1 & 0 & -1 & 1 & 1 & 0 & -1 & 1 & 0 & 1 \end{pmatrix}$$

$$\text{Event B: } \begin{pmatrix} -1 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & 1 & 1 \end{pmatrix}$$

Memory Traces Encoded with L set to .5

$$\text{Trace A: } \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \end{pmatrix}$$

$$\text{Trace B: } \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Retrieval with Event A as a Test Candidate

$$\text{Probe: } \begin{pmatrix} 1 & 0 & -1 & 1 & 1 & 0 & -1 & 1 & 0 & 1 \end{pmatrix}$$

$$\text{Trace A: } \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \end{pmatrix}$$

$$\text{Activation} = .187$$

$$\text{Trace B: } \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\text{Activation} = .001$$

$$\text{Echo Intensity} = .187 + .001 = .188$$

FIGURE 1. Example of the steps involved in simulating memory performance with MINERVA 2.

added to derive the echo intensity, which represents the amount of evidence that Event A was experienced.

In the MINERVA 2 model for this experiment, four male and four female faces were each represented as a vector of ten feature values. The faces within each gender had both shared and unique features. For each shared feature, a value of 1, 0, or -1 was randomly selected and assigned to every face within a gender. The shared features for male and female faces were selected independently, so the shared features had different values for the different genders. Each unique feature for all eight faces was randomly selected from the available feature values. Unique features were selected independently for all faces both between and within genders. In the low overlap condition, the faces within a gender had two shared features and eight unique features. In the high overlap condition, the faces within a gender had eight shared features and two unique features. Vectors representing the “average” face in each gender set were constructed by assigning each feature in the average vector to the most common feature value among the four faces in the set. If two feature values were equally common, one of them was selected at random for the average face.

Figure 2 displays example vectors representing faces in the low- and high-overlap conditions as well as an “average” face that summarizes the features of the set into a single representation. For the low-overlap faces, the first two features are shared features that have the same value for every face in the set. The values of the remaining eight features vary from face to face. The features in the average face are set to the most common values across the four faces in the set. The seventh feature has a value of zero for two faces and a value of 1 for two faces, and one of these two values was selected at random to appear in the average representation. As a result of the heterogeneity of the low-overlap faces in the set, the average face is not a close

Low-Overlap Condition

$$\begin{array}{l} \text{Face 1} \quad \left[\begin{array}{cccccccccc} 1 & -1 & 0 & 1 & 1 & 0 & -1 & 1 & 0 & 1 \end{array} \right] \\ \text{Face 2} \quad \left[\begin{array}{cccccccccc} 1 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \end{array} \right] \\ \text{Face 3} \quad \left[\begin{array}{cccccccccc} 1 & -1 & 0 & 0 & -1 & 1 & -1 & -1 & 1 & 0 \end{array} \right] \\ \text{Face 4} \quad \left[\begin{array}{cccccccccc} 1 & -1 & -1 & -1 & 0 & -1 & 1 & 0 & 1 & -1 \end{array} \right] \\ \text{Average} \quad \left[\begin{array}{cccccccccc} 1 & -1 & 0 & 0 & -1 & 1 & -1 & 0 & 1 & 1 \end{array} \right] \end{array}$$

High-Overlap Condition

$$\begin{array}{l} \text{Face 1} \quad \left[\begin{array}{cccccccccc} -1 & 0 & 1 & 1 & 0 & -1 & 1 & -1 & 1 & -1 \end{array} \right] \\ \text{Face 2} \quad \left[\begin{array}{cccccccccc} -1 & 0 & 1 & 1 & 0 & -1 & 1 & -1 & 1 & 0 \end{array} \right] \\ \text{Face 3} \quad \left[\begin{array}{cccccccccc} -1 & 0 & 1 & 1 & 0 & -1 & 1 & -1 & 0 & 1 \end{array} \right] \\ \text{Face 4} \quad \left[\begin{array}{cccccccccc} -1 & 0 & 1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 \end{array} \right] \\ \text{Average} \quad \left[\begin{array}{cccccccccc} -1 & 0 & 1 & 1 & 0 & -1 & 1 & -1 & 1 & -1 \end{array} \right] \end{array}$$

FIGURE 2. Example of individual face vectors and average vectors created by the MINERVA 2 model in the low- and high-overlap conditions.

match to any of the individual faces. For example, comparing the feature values in the average vector and in the Face 4 vector reveals only four matching features. For the high-overlap faces, the first eight feature values are shared, and only the last two features have varying values from face to face. Thus, these faces are highly homogeneous, resulting in a high degree of feature overlap between the average face and each individual face. For example, nine of the ten features in the average vector match the features in the Face 4 vector.

Item vectors were generated by randomly assigning each of 20 item features to one of the available feature values, and a total of 24 vectors were independently generated. Each face vector was paired with three of the item vectors. Merging the face and item vectors created event vectors with 30 total features (10 face features and 20 item features). These 30-feature vectors were encoded into memory traces with L set to .45. The number of features in each trace and the learning rate were selected so that the simulations would run in a reasonable amount of time and produce a level of performance in the same general range that is expected for the experiments. Changing these parameters affects the overall level of performance in each of the conditions but does not profoundly impact the qualitative pattern across conditions produced by the model. The number of event features and the value of L fall in the range of values used in previous applications of the model (e.g., Hintzman, 1986; Hintzman, 1988).

At retrieval, memory probes composed of face and item features were matched against all encoded items to derive echo intensity values. A male probe was created by combining male facial features with the item features of the test word, and this probe was matched against memory to derive a male intensity value. A female probe was created by combining female facial features with the item features of the test word, and this probe was matched against memory to derive a female intensity value. Source decisions were based on an intensity-

difference variable computed by subtracting the female intensity value from the male intensity value. Thus, the model assumes that participants select among sources by comparing the amount of evidence that the test item was studied with a male face to the amount of evidence that the test item was studied with a female face. Basing source decisions on the intensity-difference variable results in a decision space that is represented in Figure 3, which shows evidence distributions for both male and female items. Intensity difference values for female items are approximately normally distributed with a negative mean. The mean is negative because the intensity difference is computed by subtracting the female intensity from the male intensity, and female items will tend to match memory probes containing female features better than memory probes containing male features. Intensity difference values for male items have a positive mean, because male intensities tend to be higher than female intensities for male items. To select responses, participants must set response criteria along the intensity difference continuum. Figure 3 displays three criteria partitioning the continuum into four response regions, and the response associated with each region is identified in quotation marks. The ability to discriminate sources is determined by the overlap in the male and female distributions, and this overlap is jointly determined by the distance between the distribution means and the variability of each distribution.

The cueing conditions determined which male features were reinstated in the male probe and which female features were reinstated in the female probe. In the uncued condition, the model reinstated the features of the average faces within each gender along with the item features of the test word. When a particular face was externally reinstated as a retrieval cue, the features of the cued face were used in the probe matching the gender of the cue. For example, when a female face was reinstated at test, the features of this face were used in the female probe.

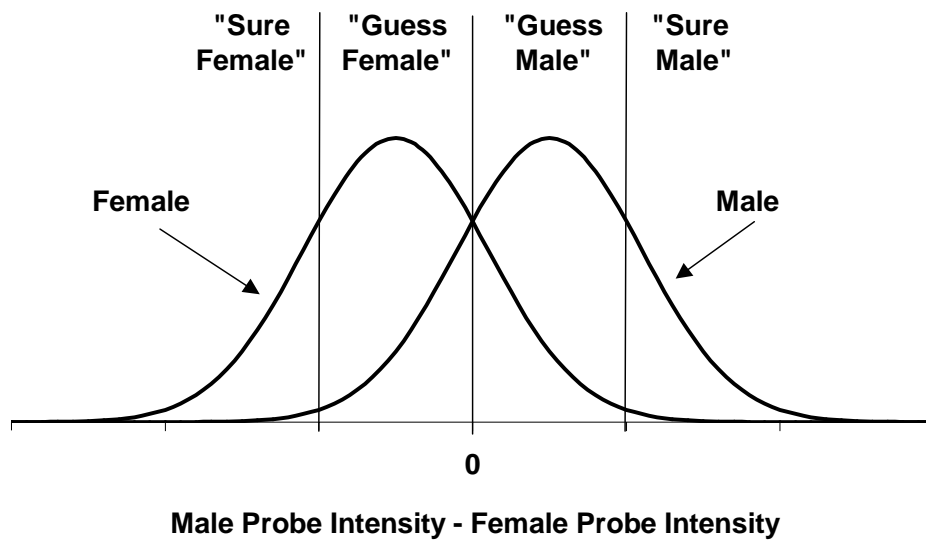


FIGURE 3. Representation of the decision space involved in discriminating male from female items based on global activation values. The y-axis represents the intensity-difference variable created by subtracting the echo intensity of the female probe from the echo intensity of the male probe for each test item. The vertical lines represent response criteria, and the words in quotations give the response associated with each decision region.

The average face was reinstated in the probe mismatching the gender of the cue. For example, when a female face was reinstated at test, the average male face was used in the male probe.

For the three items paired with each face at encoding, one was tested in the uncued condition, one was tested in the male-cued condition, and one was tested in the female-cued condition. When the face presented at test matched the item's prior presentation (e.g., a female cue for an item encoded with a female face), the facial features reinstated at test were always those of the particular face paired with the item at encoding. When the face presented at test mismatched the item's prior presentation (e.g., a female cue for an item encoded with a male face), one of the faces in the cued gender was selected, and the particular faces were rotated so that each male and female face appeared an equal number of times as a mismatched cue.

Experiment 1 Simulation Results

Memory performance measures in all experiments were derived from signal detection theory (SDT; Macmillan & Creelman, 2005). SDT assumes that memory decisions are based on a single, continuous variable representing the amount of evidence in favor of a memory candidate. SDT parameters measure the position and standard deviation of the target distribution (the distribution on the right) relative to the lure distribution (the distribution on the left). The lure distribution is assumed to have a mean of 0 and a standard deviation of 1. The target distribution has a mean of μ and a standard deviation of σ . Because the lure distribution has a mean of 0, μ can be alternatively defined as the mean of the target distribution or as the distance between the target and lure distributions. Both μ and σ are measured in terms of the lure distribution's standard deviation. For example, $\mu = 1$ indicates that the target distribution is 1 standard deviation above the lure distribution and $\sigma = 2$ indicates that the target distribution has a standard deviation that is twice the value of the lure distribution's standard deviation. In the

simulation, the mean and standard deviation of the male and female distributions were directly computed from the intensity-difference values, and SDT parameters were calculated directly from these means and standard deviations.

Table 1 displays the memory parameters output by the model. The critical predictions of the model concern the effects of retrieval cues on source memory, and source memory is a joint function of the distance between the male and female distributions and the variability of the distributions. Again, the evidence variable represented by the distributions is the difference in the intensity values produced by male and female probes (as in Figure 3), corresponding to the assumption that participants compare the amount of evidence that each test candidate was studied with a male versus a female face. In Table 1, μ was computed by subtracting the mean of the female (lure) distribution from the mean of the male (target) distribution and dividing by the standard deviation of the female distribution. Thus, μ is the distance between the distributions measured in terms of the female distribution's standard deviation. The σ column in Table 1 is the standard deviation of the male distribution divided by the standard deviation of the female distribution; therefore, a value of 1 indicates that the male and female distributions have the same standard deviation, values greater than 1 indicate that the male distribution is more variable than the female distribution, and values less than 1 indicate that the male distribution is less variable than the female distribution. Thus, the μ and σ values seen in Table 1 are adjusted to correspond to the SDT assumption that the lure (female) distribution has a mean of zero and a standard deviation of 1. The table also shows A_z , a statistic that is computed from the μ and σ values to combine these parameters into a single memory performance measure. A_z is analogous to a proportion of correct responses, and it varies from .5 when performance is at chance to 1 when performance is perfect.

TABLE 1

Memory parameters predicted by the MINERVA 2 model for Experiment 1

Cueing Condition	Overlap Condition and Memory Parameter					
	Low Overlap			High Overlap		
	μ	σ	A_z	μ	σ	A_z
Uncued	.98	.99	.76	1.43	1.01	.84
Male-Cued	1.62	1.24	.85	1.56	1.04	.86
Female-Cued	1.31	.82	.85	1.52	.95	.86

For the low overlap group in Table 1, μ values are higher in the two conditions where external cues are provided than in the uncued condition. Additionally, the male-cued μ is higher than the female-cued μ . The reason for this latter difference is that, when cues are presented, the variability of the distribution that matches the gender of the cue is higher than the variability of the other distribution. This can be seen in the σ values across conditions: σ is greater than one in the male-cued condition (1.24) and less than one in the female-cued condition (.82). Combining the μ and σ values into A_z reveals that source memory improves when external cues are provided relative to the uncued condition, and that source memory in the male-cued and female-cued conditions is equivalent.

In the low overlap group, cues increase source discriminability because they create a better match between the facial features in the memory probe and the facial features encoded for the test item. With no external cue, the model reinstates the “average” face for a given gender, and many of the features in this average face mismatch the individual faces. When a correct-gender external cue is provided, the model reinstates the exact facial features encoded with the

test item. Therefore, female cues increase the activation of the female probe for items studied with a female face, and male cues increase the activation of the male probe for items studied with a male face.

Results for the high-overlap group in Table 1 show that external cues had a much smaller effect on source memory in this condition. The A_z values for the cued conditions are only slightly higher than the uncued condition. Cueing effects are attenuated because internal reinstatement is almost as effective as external reinstatement in this condition. The faces within each gender share most of their features in this condition, so the average face used on uncued trials is an extremely good match to each individual face. When a matching external cue is provided, the cue provides only a slightly better match than the cue that would have been internally reinstated.

Table 2 shows the proportion of items in each source claimed to have been studied with a male face across all cells of the design. These values were obtained from the model results by setting a criterion at a value of zero on the intensity-difference continuum (this is the center criterion in Figure 3, page 32). The model claimed that the test item was studied with a male face for any intensity-difference value above the criterion. The table also shows d' values calculated using the proportion of male items called male as the hit rate and the proportion of female items called male as the false alarm rate. These d' scores are a measure of source discriminability much like A_z , but the d' scores are calculated directly from the hit and false alarm rates as opposed to being calculated from the evidence values generated by the model. If the male and female evidence distributions are assumed to have equal variances, d' measures the distance between the distributions (so, $d' = \mu$ when the male and female distributions have equal variances). The d' values in Table 2 clearly reveal the beneficial effects of retrieval cues in the

low-overlap condition: relative to the uncued condition, d' is higher with both male and female cues. In the high-overlap condition, d' scores are very similar across cueing conditions. Importantly, the hit and false alarm rate values in Table 2 were calculated based on the assumption that response bias does not vary across cueing conditions (the response criterion was always set at a value of zero). With no bias effects, the hit and false alarm rates show a simple pattern of cueing effects in the low-overlap condition: male cues improve performance for male items (i.e., they increase the hit rate) and female cues improve performance for female items (i.e., they decrease the false alarm rate). Previous studies suggest that retrieval cues do influence response bias (e.g., Dodson & Shimamura, 2000), so the pattern of hit and false alarm rates presented in Table 2 may not match the hit and false alarm rates observed in the first experiment. For this reason, I will focus on d' scores when evaluating the success of the model's predictions, because this measure corrects for differential bias across conditions.

TABLE 2

MINERVA 2 predictions for the source memory hit rates, false alarm rates, and d' scores across the overlap and cueing conditions of Experiment 1

Cueing Condition	Overlap Condition and Source Memory Measure					
	Low Overlap			High Overlap		
	HR	FAR	d'	HR	FAR	d'
Uncued	.68	.30	.99	.77	.24	1.46
Male-Cued	.81	.29	1.43	.78	.23	1.51
Female-Cued	.70	.19	1.40	.78	.22	1.54

Note: HR = hit rate (the proportion of male items called “male”); FAR = false alarm rate (the proportion of female items called “male”).

An Alternative Internal Reinstatement Strategy. The preceding model simulation was based on the assumption that participants construct a summary or “average” face for each gender and use this summary representation as an internal face cue when faces are not externally reinstated. This is a reasonable strategy given that face information must be included in retrieval probes to make source decisions, and the multitude of studied faces should prevent the internal reinstatement of individual faces. However, participants could potentially use another strategy for internally reinstating faces. Namely, on uncued trials, they may just randomly select one of the male faces and one of the female faces to construct their male and female retrieval probes. I conducted a second simulation to ensure that the model predictions do not change if this alternative strategy is assumed. The second model matched the first in all respects except the male and female features in the retrieval probes for uncued items. Instead of the average male and female faces, on each test trial one male and one female face was selected at random and the features of these faces were included in the retrieval probe. The alternative model produced the same qualitative pattern reported for the first simulation. Specifically, the cueing variable had little effect on source memory in the high overlap condition ($A_z = .84, .86, \text{ and } .86$ in the uncued, male-cued, and female-cued conditions, respectively). In the low overlap condition, source memory was poor with no external cues ($A_z = .68$), but increased substantially when male ($A_z = .82$) or female ($A_z = .82$) cues were provided.

Experiment 1 Method

Participants. One hundred fifty-six Louisiana State University undergraduates were randomly assigned to the low- or high-overlap conditions in equal numbers. Participants received extra credit in their psychology courses for volunteering.

Power Analysis. I used G-Power (Erdfelder, Faul, & Buchner, 1996) to calculate the power achieved in each of the experiments. The critical prediction of the simulation model is an interaction between the overlap and cueing variables, so I evaluated the power to detect this interaction. With one between-subjects and one within-subjects variable, the non-centrality parameter (δ) for the interaction effect is calculated by the formula

$$\delta = N \times m \times f^2 / (1 - \rho)$$

where N is the total number of participants, m is the number of levels for the within-subjects factor, f^2 is the between-subjects effect size, and ρ is the correlation between the levels of the within-subjects factor (Buchner, Erdfelder, & Faul, 1997). f^2 is the ratio of the variance of group means to the variance of the scores within the groups (Cohen, 1988). Power analyses for all experiments used $f^2 = .02$, which can be considered a small effect size (Erdfelder, Faul, & Buchner, 1996). To select an appropriate value for ρ , I calculated the correlation (Pearson's r) between the levels of within-subject factors in several source and recognition memory experiments previously conducted in our laboratory. These experiments included a study evaluating the effects on recognition performance of reinstating the font size of studied words, as well as a study evaluating the effects of voice reinstatement on source memory. Correlations were consistently near a value of .4, so I used this value in all power analyses. I calculated the non-centrality parameter for this experiment and entered this value into G-Power (Erdfelder, Faul, & Buchner, 1996) to perform the power analysis. With 156 participants, the power to detect a small interaction effect was .95 in Experiment 1.

Design. This study conformed to a 2 (overlap group) x 3 (cueing condition) factorial design. Overlap group was manipulated between subjects, and the levels of this variable were low overlap (faces within a gender had few shared and many unique features) and high overlap

(faces within a gender had many shared and few unique features). Cueing condition was manipulated within subjects, and the levels of this variable were uncued (no face seen at test), cued with a male face, or cued with a female face.

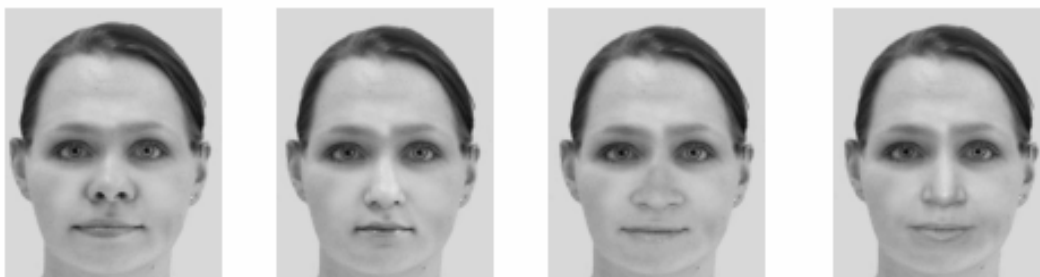
Materials. Ninety-six words were selected from the MRC psycholinguistic database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). All words were between 5 and 8 letters in length, had familiarity ratings between 500 and 700 on a scale ranging from 100 to 700, and had concreteness ratings between 300 and 500 on a scale ranging from 100 to 700. Male and female faces were constructed using the morphases website (<http://www.morphases.com/editor/>). The morphases software blends elements of pictures of different faces into a single composite face, and the program allows users to vary the hair, forehead, eyebrows, eyes, nose, mouth, ears, and chin of the composite face. To construct the faces for this experiment, I began with a base female face and a base male face. The stimuli for each condition were created by exchanging features of the base face with features from another person of the same gender. For the low-overlap condition, the four stimulus faces in each gender were produced by replacing the eyebrows, eyes, nose, mouth, ears, and chin of each base with the eyebrows, eyes, nose, mouth, ears, and chin of three different people of the same gender as the base. For the high-overlap condition, only the nose and mouth of each base were replaced. The same pictures were used to exchange features in the high and low overlap conditions. Figure 4 displays the male and female faces that were used in each overlap condition. As depicted in Figure 4, faces appeared in black-and-white in the experiment.

At study, each face in both genders was paired with 12 words. Four of the words studied with each face appeared with no face at test, four appeared with a male face, and four appeared with a female face. When the gender of the test face matched the gender paired with the word at

Female – Low Overlap



Female – High Overlap



Male – Low Overlap



Male – High Overlap



FIGURE 4. Female and male faces that were used for the low- and high-overlap conditions.

encoding (e.g., a male cue for a male item), the exact face paired with the word at encoding appeared at test. When the gender of the test face mismatched the gender paired with the word at encoding, one of the faces from the opposite gender was randomly selected to appear at test under the constraint that all faces appeared an equal number of times.

Procedure. Before the study phase, participants were told that they would see words paired with faces, and that their memory for the words would be tested later in the experiment. They were told to imagine the person in the picture saying the presented word. Following the instructions, the word-face pairs on the study list appeared in random order, and each remained on the screen for 3 s. The pictures appeared in the center of the computer screen, and words appeared below the picture. Next, participants performed a distracter task in which they rearranged the segments of a scrambled picture on the computer screen until each segment was in its appropriate position. None of the pictures in the distracter task depicted human faces. Participants completed as many puzzles as they could solve in three minutes, and then they heard the test instructions. The test instructions informed participants that they would see words that appeared in the study phase, and that they must decide whether each word appeared with a male or a female face. They were told to press 1 if they were sure the word was seen with a female picture, press 2 if they guessed the word was seen with a female picture, press 3 if they guessed the word was seen with a male picture, and press 4 if they were sure the word was seen with a male picture. Participants were also told that they would sometimes see a face along with the word at test, but that the test face would not always be the face that was paired with the word in the study phase. The test words appeared in random order and, on cued trials, the test face appeared above the word as in the study phase.

Experiment 1 Results and Discussion

Table 3 displays the hit rates, false alarm rates, and d' scores from the uncued, male cued, and female cued conditions. Again, the hit rate is the proportion of male items called “male” (i.e., receiving a rating of 3 or 4) and the false alarm rate is the proportion of female items called “male.” The table reveals that the cueing manipulation influenced participants’ general bias to say “male.” With male cues, participants made more “male” responses to both male and female items; that is, both the hit rate and the false alarm rate increased relative to the uncued condition. With female cues, participants made fewer “male” responses to both male and female items; that is, both the hit rate and the false alarm rate decreased relative to the uncued condition. Thus, participants showed a bias to respond with the gender of the face they saw at test. This pattern across cueing condition was observed for both the low- and high-overlap groups. The d' scores correct for differential bias to call test candidates “male,” so these values will be the focus of all inferential analyses. For the low-overlap group, d' appeared to be larger in the male- and female-cued conditions than in the uncued condition, but the differences were small (less than 2 standard errors). For the high-overlap group, d' scores were similar across the cueing conditions. d' scores were analyzed with a 2 (overlap condition) x 3 (cueing condition) mixed factors ANOVA, with overlap condition as a between subjects factor and cueing condition as a within subjects factor. The main effects of overlap and cueing both failed to reach significance, $F(1, 154) = .07$, ns , and $F(2, 308) = .97$, ns , respectively. The interaction of the overlap and cueing variables was also non-significant, $F(2, 308) = .61$, ns . Because the model made specific predictions regarding cueing effects in the low- and high-overlap conditions, I performed a specific contrast for each overlap group even though the initial ANOVA revealed no interaction (I continue this practice throughout the other experiments). The contrast compared the uncued condition to the male- and

female-cued conditions. The contrast was not significant for the high-overlap group, $F(1, 77) < 1$, *ns*, indicating that source memory did not differ between the uncued condition and the two conditions in which cues were provided. The contrast was also non-significant for the low-overlap group, $F(1, 77) = 1.90$, *ns*.

TABLE 3

Observed source memory hit rates, false alarm rates, and d' scores across the overlap and cueing conditions of Experiment 1

Cueing Condition	Overlap Condition and Source Memory Measure					
	Low Overlap			High Overlap		
	HR	FAR	d'	HR	FAR	d'
Uncued	.62 (.02)	.42 (.02)	.50 (.06)	.64 (.02)	.43 (.02)	.55 (.06)
Male-Cued	.74 (.01)	.52 (.02)	.58 (.07)	.74 (.02)	.57 (.02)	.51 (.07)
Female-Cued	.50 (.02)	.28 (.02)	.61 (.06)	.52 (.02)	.31 (.02)	.58 (.06)

Note: HR = hit rate (the proportion of male items called “male”); FAR = false alarm rate (the proportion of female items called “male”). Standard errors are in parentheses.

For a second method of analysis, the rating scale response frequencies were used to fit a signal detection model. Within each overlap group, the empirical data included a total of 24 response frequencies; that is, the frequency with which participants responded to the two sources of items with each of the four rating responses in the three cueing conditions ($2 \times 4 \times 3 = 24$). For each set of four rating responses, one of the response frequencies is not independent (i.e., one frequency is fixed once the values of the other three frequencies and the total frequency are known). Thus, there are 18 *independent* response frequencies in the dataset for this experiment.

The signal detection model had five parameters in each condition: μ – the distance between the male and female distributions; σ – the standard deviation of the male distribution; λ_1 – the position of the criterion separating “sure female” from “guess female” responses; λ_2 – the position of the criterion separating “guess female” from “guess male” responses; and λ_3 – the position of the criterion separating “guess male” from “sure male” responses (the λ parameters correspond to the vertical lines in Figure 3, page 32). The statistic A_z was calculated using the μ and σ parameters from each condition. When the model is fit to data from the three cueing conditions, there are a total of 15 parameters. This 15 parameter model was separately fit to the low- and high-overlap groups via maximum likelihood estimation, and model fit was evaluated using the G^2 statistic. This statistic is a measure of the overall deviation of the empirical response frequencies from the frequencies predicted by the model, with smaller values indicating a better fit. G^2 values are approximately χ^2 distributed, and significance values for all hypothesis tests were computed from a χ^2 distribution. Cueing effects were tested by evaluating the difference in fit between the full 15 parameter model and a reduced model in which the discriminability parameters (μ and σ) were constrained to be equal across the cueing conditions (this eliminates 4 parameters because a single μ replaces the 3 μ parameters of the full model and a single σ replaces the three σ parameters of the full model). This is a χ^2 test with 4 degrees of freedom, and the statistic is simply the difference in fit between the full and reduced models.

Table 4 displays the parameters produced by the signal detection model. The most important data are found in the A_z column, as this is the model-based measure of source discriminability. The A_z values show little effect of retrieval cues on source memory in either the low or high overlap conditions. Constraining memory parameters (i.e., μ and σ) to be equal across the cueing conditions did not significantly change the fit of the model for the high-overlap

group, $G^2(4) = 3.67$, *ns*. For the low-overlap group, assuming equal source memory across cueing conditions significantly worsened model fit, $G^2(4) = 18.85$, $p < .001$. However, performance in the male-cued condition ($Az = .67$) was no higher than the uncued condition ($Az = .67$), and the female-cued condition ($Az = .69$) was only slightly higher than the uncued condition. Although the cueing effect was significant in the inferential test, the specific parameter values reveal little evidence for the prediction that retrieval cues improve source memory.

TABLE 4

Memory parameters generated by the signal detection models for the low- and high-overlap groups in Experiment 1

Cueing Condition	Overlap Condition and Memory Parameter					
	Low Overlap			High Overlap		
	μ	σ	Az	μ	σ	Az
Uncued	.61	1.02	.67	.59	.97	.66
Male-Cued	.58	.87	.67	.53	.99	.65
Female-Cued	.72	1.10	.69	.64	1.04	.67

The pattern of σ parameters observed in the low-overlap condition also appears to violate the simulation model's predictions. The low-overlap data in Table 1 (page 23) reveal that, in the simulation results, σ was near 1 in the uncued condition, increased in value when male cues were provided, and decreased in value when female cues were provided. Table 4 shows a divergent pattern in the fits to the empirical data: σ was near 1 in the uncued condition, *decreased* in value when male cues were provided, and *increased* in value when female cues were provided. The

reason for this discrepancy is unclear. One strong possibility is that retrieval cues actually had no systematic influence on the σ parameter, just as cues did not appear to influence overall source performance. The pattern of σ values in Table 4 may simply reflect random variation.

The results of Experiment 1 do not show the strong interaction predicted by the MINERA 2 model and by the cue discrepancy hypothesis. In the simulation, retrieval cues had little effect in the high-overlap condition, but dramatically improved source memory in the low-overlap condition. In the empirical results, cues did not substantially impact memory performance for either overlap group. The d' scores were not significantly affected by cues, and the SDT model also yielded little evidence that cues improve source memory. Following the second experiment, I will discuss possible reasons why the results were inconsistent with the predictions.

EXPERIMENT 2

The next experiment explored the effects of cue reinstatement on recognition memory. Participants studied words paired with one of four faces from a single gender, and, as in the previous experiment, there was either high or low overlap in features among the faces. At test, participants were asked to discriminate studied words from words that did not appear on the study list, and a studied face was reinstated as a cue on some of the trials. For studied items that were cued at test, sometimes the face seen with the item at encoding was reinstated (match trials) and sometimes one of the other faces was reinstated (mismatch trials).

Simulation Model for Recognition Memory in Experiment 2

One set of four faces was used in this simulation (i.e., one gender), and the faces in the high and low overlap conditions were generated in an identical fashion to the first simulation. An “average” representation of these four faces was created as in the first simulation. Forty-eight vectors of 20 item features were randomly generated. Each face was paired with six separate item vectors. These items were considered targets, and they were encoded into memory with L set to .45. The other 24 item vectors were not encoded into the memory set, but were used as lures at test.

At test, a retrieval probe consisting of face and item features was matched against memory for each of the 24 targets and 24 lures, and decisions were based on the echo intensity produced by the probe. The specific facial features reinstated at test depended on the cueing condition. Of the six items paired with each face at encoding, two were uncued at test, two were cued with a matching face, and two were cued with a mismatching face. Additionally, 16 lures were cued with a studied face and 8 were uncued. For uncued targets and lures, the average face was reinstated in the retrieval probe along with the item features of the test word. For targets in

the match condition, the facial features of the face seen with the test word at encoding were reinstated. For targets in the mismatch condition, one of the faces that were not paired with the test item was randomly selected and reinstated in the test probe. For cued lures, one of the faces from the study phase was reinstated, and the faces were rotated such that each appeared an equal number of times.

Experiment 2 Simulation Results

Table 5 displays the memory parameters generated by the model in each cueing condition. For the low-overlap group, matching cues increased the distance between the target and lure distributions (μ) compared to both the uncued and mismatch conditions, and the standard deviation of the target distribution (σ) showed the same pattern of variation. Most importantly, combining these parameters into the A_z measure demonstrates that recognition memory was highest in the match condition, lowest in the mismatch condition, and intermediate in the uncued condition. Therefore, matching cues facilitated recognition memory in the low-overlap condition, the same pattern seen for source memory in the first simulation. Mismatching cues impair recognition in relation to the uncued condition because a mismatching individual face does not match the features of the studied face as closely as the average face reinstated on uncued trials. In the high-overlap condition, the effects of cueing are dramatically attenuated. Therefore, the simulation model again predicts an interaction between the overlap and cueing conditions.

Table 6 displays hit and false alarm rates for each condition generated by placing a criterion one standard deviation above the mean of the lure distribution and considering all evidence values exceeding this criterion to be studied items. The same criterion was used for all conditions, so all of the false alarm rates are the same. The table also shows d' scores calculated

from the hit and false alarm rates. The d' scores reveal the same interaction as the A_z values in Table 4. With low feature overlap, matching cues substantially improve and mismatching cues substantially impair recognition. With high feature overlap, retrieval cues have a smaller effect.

TABLE 5

Memory parameters predicted by the MINERVA 2 model for Experiment 2

Cueing Condition	Overlap Condition and Memory Parameter					
	Low Overlap			High Overlap		
	μ	σ	A_z	μ	σ	A_z
Uncued	1.55	1.71	.78	1.36	1.45	.78
Match	2.34	2.02	.85	1.60	1.56	.81
Mismatch	.98	1.39	.72	1.33	1.45	.77

TABLE 6

MINERVA 2 predictions for the recognition memory hit rates, false alarm rates, and d' scores across the overlap and cueing conditions of Experiment 2

Cueing Condition	Overlap Condition and Recognition Measure					
	Low Overlap			High Overlap		
	HR	FAR	d'	HR	FAR	d'
Uncued	.63	.16	1.32	.60	.16	1.25
Match	.75	.16	1.66	.65	.16	1.38
Mismatch	.49	.16	.98	.59	.16	1.23

Note: HR = hit rate; FAR = false alarm rate.

The recognition results are similar to the source monitoring results because cueing effects in recognition are based on the same mechanisms as cueing effects in source memory. In the low-overlap group, the average face is not a close match to any of the individual faces. When the average face is reinstated on uncued trials, the facial features in the test probe are not a good match for the facial features encoded for the test word. For cued targets, the test probe contains the features of the individual face encoded for the test word, leading to an increase in memory performance. The features in a mismatching face cue are an even worse match to the studied face than the average face, leading to a drop in performance. In the high-overlap group, the average face very closely matches the features of each individual face. Therefore, providing an external cue only slightly increases the match of the facial features in the test probe to the facial features encoded for the test word. Moreover, because all of the faces are similar, mismatching cues are not as detrimental to source performance.

Alternative Internal Reinstatement Strategy. As in the first experiment, I conducted an additional simulation based on a strategy of reinstating a randomly selected face for uncued trials (as opposed to reinstating the average face). Results in the low-overlap condition showed the same pattern as the initial Experiment 2 simulation. Specifically, relative to the uncued condition ($A_z = .75$), recognition memory increased with matching cues ($A_z = .84$) and decreased with mismatching cues ($A_z = .71$). Memory performance was higher in the uncued than the mismatch condition because the random face selected on uncued trials sometimes matched the studied face, but the face reinstated on mismatch trials never matched the studied face. Also mirroring the initial simulation, cueing effects were much smaller in the high-overlap condition ($A_z = .77, .80, \text{ and } .77$ in the uncued, match, and mismatch conditions, respectively).

Experiment 2 Method

Participants. One hundred forty-four Louisiana State University undergraduates participated in exchange for extra credit in psychology courses. Participants were randomly assigned to the overlap groups in equal numbers. With $f^2 = .02$ and $\rho = .4$, the power to detect a small interaction effect was .93.

Design. Experiment 2 followed a 2 (overlap condition) x 3 (cueing condition) factorial design. The levels of the overlap variable were high- and low-overlap. The levels of the cueing variable were uncued, match, and mismatch. As in the first experiment, overlap condition was manipulated between subjects and cueing condition was manipulated within subjects.

Materials. The words and the male faces from Experiment 1 were used in this experiment. The study list consisted of 96 word-face pairs, with 24 words paired with each of the 4 male faces. The test list contained 48 studied words (12 of the words paired with each face) and 48 lures. Twenty-four targets appeared with the male face paired with the word at study (matching cues), 12 targets appeared with a face that was not seen with the target at study (mismatching cues), and 12 targets appeared alone on the screen (uncued). For mismatching target trials, one of the faces other than the face seen at encoding was randomly selected under the constraint that each face appeared an equal number of times. Similarly, half of the lures appeared with one of the faces seen at encoding, and half of the lures appeared without a face. Each face was randomly paired with six lures.

Procedure. Procedural details matched the first experiment with the exception of the test procedure. Instructions informed participants that they would see words that either did or did not appear in the study phase, and they should indicate whether or not they studied each word. Participants were told to press 1 if they were sure that the item was new (unstudied), press 2 if

they guessed that the item was new, press 3 if they guessed that the item was old (studied), and to press 4 if they were sure that the item was old. Participants were told that they would see faces on some of the trials, but that they should base their decisions only on the word that they saw because studied faces would sometimes be paired with unstudied words.

Experiment 2 Results and Discussion

Table 7 displays the hit rates, false alarm rates, and d' scores across the cueing conditions for each overlap group. The retrieval cues influenced participants' bias to claim that test candidates were "old." For both overlap groups, hit rates were higher when faces were reinstated at test than in the uncued condition regardless of whether the test face matched or mismatched the face studied with the target. False alarm rates were also higher for lures that were cued with studied faces than for uncued lures. When these bias differences are corrected with d' scores, retrieval cues appear to have had little or no effect on recognition memory. Regardless of the feature overlap in the studied faces, matching retrieval cues conferred no memory advantage compared to the uncued and mismatch conditions. The d' scores were submitted to a 2 (overlap group) x 3 (cueing condition) mixed factors ANOVA, with overlap group as a between-subjects factor and cueing condition as a within-subjects factor. The main effects of both the overlap and cueing conditions were not significant, $F(1, 142) = .02, ns$, and $F(2, 284) = .60, ns$, respectively. The interaction also failed to reach significance, $F(2, 284) = .23, ns$. Specific contrasts in the high-overlap group showed that the match condition did not significantly differ from either the uncued or mismatch conditions, $t(71) = .55, ns$, and $t(71) = .04, ns$, respectively. For the low-overlap group, the match condition was also statistically indistinguishable from the uncued and mismatch conditions, $t(71) = .17, ns$, and $t(71) = 1.15, ns$, respectively.

TABLE 7

Observed recognition memory hit rates, false alarm rates, and d' scores across the overlap and cueing conditions of Experiment 2

Cueing Condition	Overlap Condition and Recognition Measure					
	Low Overlap			High Overlap		
	HR	FAR	d'	HR	FAR	d'
Uncued	.65 (.02)	.29 (.02)	.97 (.07)	.64 (.02)	.29 (.02)	.95 (.07)
Match	.72 (.02)	.38 (.02)	.95 (.07)	.71 (.02)	.37 (.02)	.91 (.06)
Mismatch	.70 (.02)	.38 (.02)	.89 (.09)	.70 (.02)	.37 (.02)	.91 (.08)

Note: HR = hit rate; FAR = false alarm rate. The false alarm rate in the match and mismatch rows are the same value (i.e., false alarms for lures cued with a studied face). Standard errors are in parentheses.

The results of the signal detection model fit to the rating scale response frequencies from Experiment 2 appear in Table 8. For both overlap groups, the A_z values revealed no evidence that matching cues enhanced source memory performance relative to the uncued and mismatch conditions. Equating memory performance across cueing conditions did not significantly affect model fit for either the low- or high-overlap groups, $G^2(4) = 7.85, ns$, and $G^2(4) = 4.20, ns$, respectively.

As in the first experiment, the empirical results are inconsistent with the predictions of the simulation model. The d' scores showed no hint of an interaction between the overlap and cueing variables; instead, retrieval cues did not affect memory for either overlap group. The results of the signal detection model also show that cues were ineffective regardless of whether the studied faces had low or high overlap in their features.

TABLE 8

Memory parameters produced by the signal detection models for the low- and high-overlap groups in Experiment 2

Cueing Condition	Overlap Condition and Memory Parameter					
	Low Overlap			High Overlap		
	μ	σ	Az	μ	σ	Az
Uncued	1.08	1.39	.74	1.07	1.35	.74
Match	.94	1.24	.72	.97	1.21	.73
Mismatch	.86	1.27	.70	.95	1.17	.73

Why did the results of the first two experiments fail to support the prediction that cues should substantially enhance memory in the low-overlap condition but not in the high-overlap condition? One explanation is that the overlap variable failed to successfully manipulate the effectiveness of internal reinstatement. Notably, the overlap variable did not influence discriminability in Experiments 1 and 2. If a low degree of feature overlap truly decreased the effectiveness of internal reinstatement (as predicted by the MINERVA 2 model), then source performance should have been lower in the low-overlap condition than in the high-overlap condition (at least for uncued trials). Perhaps the faces used in the low-overlap condition were not sufficiently distinct to prevent effective internal reinstatement of an “average” face for each gender. Alternatively, four faces in each gender may not have been enough to prevent participants from internally reinstating each individual face at test. If participants considered each face individually, then the degree of feature overlap amongst the faces should not have impacted performance.

The specific predictions of the MINERVA 2 simulation model were not borne out by the first two experiments. Specifically, the simulations predicted that the degree of feature overlap would impact the effectiveness of internal reinstatement, and the empirical results show no evidence of such an effect. Without a successful manipulation of the effectiveness of internal reinstatement, Experiments 1 and 2 do not provide an adequate test of the cue discrepancy hypothesis. In the following experiments, I attempted a different and potentially more powerful method of disrupting internal reinstatement. Thus, these experiments provided a more valid test of the cue discrepancy hypothesis.

EXPERIMENT 3

In Experiments 3 and 4, participants studied words that were paired with either one male and one female face or with one of eight male and eight female faces. All faces within each gender were perceptually distinct, and no features were held constant across the faces. Thus, the multiple-face condition in Experiments 3 and 4 can be considered a more dramatic version of the low-overlap condition in Experiments 1 and 2. All participants completed a source test in which they decided whether each test word was studied with a male or a female face, and external cues were provided on selected test trials. Thus, both of the remaining experiments evaluate cueing effects in source memory.

Participants who encounter only a single male and female face across study trials should have no difficulty internally reinstating appropriate face cues on uncued trials: they can simply reinstate the male face from the study phase to search for male evidence and the female face from the study phase to search for female evidence. In contrast, participants who encounter eight male and eight female faces across study trials should not be able to effectively internally reinstate face information on uncued trials. Due to the heterogeneity of the faces within each gender, participants will not be able to construct an average face that will be a close match to any of the individual faces. If participants randomly select an individual face to reinstate on uncued trials, the selected face will have a low probability of matching the face actually studied with the test candidate. Moreover, the multiple-face conditions in Experiments 3 and 4 involved twice as many faces as the first two experiments, which should prevent participants from internally reinstating all of the individual faces on each test trial. In short, the number-of-faces variable in Experiments 3 and 4 should provide a more successful manipulation of the effectiveness of internal reinstatement than the overlap variable in the first two experiments.

According to the cue discrepancy hypothesis, external cues should affect memory performance only when internal reinstatement is ineffective. Therefore, this hypothesis predicts a larger cueing effect in the multiple-face condition than in the single-face condition. Moreover, memory performance in the single-face condition will likely be unaffected by the external cueing manipulation. With no ambiguity as to the male and female face potentially studied with each test candidate, participants should be able to internally reinstate all of the information available in the external cues.

Experiment 3 Method

Participants. Data were collected from 104 Louisiana State University undergraduates who received course extra credit for volunteering. Participants were randomly assigned to the single-face and multiple-face conditions in equal numbers. With $f^2 = .02$ and $\rho = .4$, the power to detect an interaction effect was .83.

Design. This study conformed to a 2 (face number) \times 3 (cueing condition) factorial design. Face number was manipulated between subjects, and the levels of this variable were multiple face (eight male and eight female faces) and single face (one male and one female face). Cueing condition was manipulated within subjects, and the levels of this variable were uncued, male-cued, and female-cued.

Materials. This experiment involved the same stimulus words as the first two experiments. Eight male and eight female faces were selected from Maner et al. (2003). Unlike the previous experiments, the faces were not altered to create overlap in features from face to face. The encoding list contained 96 words, with 48 words seen with a female face and 48 seen with a male face. The assignment of gender to individual words was counterbalanced across participants. For the single-face group, one of the male and one of the female faces was selected

at random, and these faces appeared throughout the study phase. In the multiple-face group, each of the eight male faces was randomly paired with six words, and each of the eight female faces was randomly paired with six words. The test list contained the 96 words seen at encoding. Of the 48 words studied with a male face, 16 appeared with no face at test, 16 appeared with the male face studied with the word, and 16 appeared with a female face. Of the 48 words studied with a female face, 16 appeared with no face at test, 16 appeared with a male face, and 16 appeared with the female face studied with the word. When the gender of the test cue mismatched the gender of the face studied with the test word, one face from the cued gender was selected at random under the constraint that all faces appeared an equal number of times at test.

Procedure. All procedural details matched those of Experiment 1 with three exceptions. First, the study instructions asked participants to imagine the person that they saw on each trial “demonstrating” the word they saw on each trial (as opposed Experiment 1’s instructions to imagine the person saying the word). For example, participants were told that, if the word was a verb, they could imagine the person in the picture acting out the word. Conversely, if the word was a noun, participants were told that they could imagine the person in the picture holding up or pointing to the item referred to by the word. Second, each word/picture pair remained on the screen for 5 s in the study phase (up from 3 s in Experiment 1), with 200 ms of blank screen following each word. These first two alterations to the Experiment 1 procedure were designed to improve the overall level of source memory. Third, at test, participants responded using a 6-point instead of a 4-point rating scale. The points of the rating scale were labeled “Very Sure Female,” “Sure Female,” “Guess Female,” “Guess Male,” “Sure Male,” and “Very Sure Male.”

Experiment 3 Results and Discussion

Table 9 displays the hit rates, false alarm rates, and d' scores across the cueing conditions for each face number group. Relative to the uncued condition, presenting a male cue at test led to increases in both the hit rate and the false alarm rate, indicating that male cues increased participants' bias to use the "male" response for both items studied with a male face (hit rate) and items studied with a female face (false alarm rate). With female cues, both the hit rate and the false alarm rate were lower than the uncued condition, which demonstrates that female cues decreased participants' bias to use the "male" response for both sources of items. This pattern of hit and false alarm rates is evident for both the multiple-face and the single-face groups. Thus, as in the first experiment, participants displayed a bias to respond with the gender of the face they saw at test.

TABLE 9

Observed recognition memory hit rates, false alarm rates, and d' scores across the face number and cueing conditions of Experiment 3

Cueing Condition	Face Number and Source Measure					
	16-Face			2-Face		
	HR	FAR	d'	HR	FAR	d'
Uncued	.65 (.02)	.37 (.02)	.75 (.11)	.73 (.02)	.30 (.02)	1.20 (.11)
Male	.79 (.02)	.46 (.03)	.98 (.11)	.78 (.02)	.35 (.03)	1.23 (.11)
Female	.54 (.03)	.23 (.02)	.92 (.12)	.68 (.03)	.24 (.03)	1.27 (.12)

Note: HR = hit rate (the proportion of male items called "male"); FAR = false alarm rate (the proportion of male items called "female"). Standard errors are in parentheses.

Combining the hit and false alarm rates into d' reveals the effect of cues on source discrimination. For the multiple-face group, d' scores were higher with male and female test cues than with no external cue. For the single-face group, d' scores were similar across all levels of the cueing manipulation. A 2 (Face Number) \times 3 (Cueing Condition) ANOVA on the d' scores showed a main effect of face number, $F(1, 102) = 5.50, p < .05$, reflecting poorer source discrimination in the multiple-face condition ($M = .88$) than in the single-face condition ($M = 1.23$). The main effect of the cueing variable was also significant, $F(2, 204) = 3.11, p < .05$. This main effect arose because performance was lower in the uncued condition ($M = .97$) than in the male-cued ($M = 1.10$) and female-cued ($M = 1.10$) conditions. The interaction of face number and cueing was not significant, $F(2, 204) = 1.46, ns$; however, the means in Table 9 suggest that the cueing effect was driven almost entirely by the participants in the multiple-face condition. Indeed, the contrast comparing the uncued condition to the male- and female-cued conditions showed a significant cueing effect for the multiple-face group [$F(1, 51) = 3.11, p < .01$] but not for the single-face group [$F(1, 51) = .49, ns$].

The parameters produced by the signal detection model fit to the response frequencies from Experiment 3 appear in Table 10. The A_z values show a slight advantage for the male- and female-cued conditions relative to the uncued condition in the multiple-face group, but no apparent cueing effect in the single-face group. In the model for the multiple-face participants, equating memory performance across cueing conditions did not significantly affect model fit, $G^2(4) = 4.40, ns$. Therefore, the apparent effect of retrieval cues in the A_z scores did not reach significance. Cueing condition also did not affect source memory in the single-face condition, $G^2(4) = 2.12, ns$.

TABLE 10

Memory parameters produced by the signal detection models for the multiple- and single-face groups in Experiment 3

Cueing Condition	Face Number and Memory Parameter					
	16 Faces			2 Faces		
	μ	σ	Az	μ	σ	Az
Uncued	.85	.97	.73	1.28	1.03	.81
Male-Cued	1.03	1.03	.76	1.25	.93	.82
Female-Cued	.98	1.04	.75	1.25	1.00	.81

The results support the contention that increasing the number of faces within each gender creates internal-reinstatement problems, in that source performance suffered in the multiple-face condition relative to the single-face condition. The results also reveal suggestive evidence that the effectiveness of internal reinstatement interacts with external cueing manipulations. The d' scores differed significantly across cueing conditions for the multiple face group but not for the single face group. However, the cueing effect was small even in the multiple face group, and the interaction of the face number and cueing variables failed to reach significance. Moreover, the signal detection model did not produce a significant effect of retrieval cues in the multiple-face condition. Overall, the results do not strongly support the cue discrepancy hypothesis, although a hint of the predicted interaction was observed.

EXPERIMENT 4

In the next experiment, I altered the cueing procedure in an attempt to more definitively support the cue discrepancy hypothesis. In Experiment 3, a face of the incorrect gender was displayed for half of the cued test trials; that is, half of the cued male items appeared with a female face at test and half of the cued female items appeared with a male face at test. The incorrect trials allowed me to estimate the effects of retrieval cues on response bias; however, incorrect trials may have also limited participants' use of the cues provided by the experimenter. For example, participants who realized that the cued face was often incorrect may have begun to pay less attention to the external cues. For Experiment 4, I devised a technique for presenting a correct face on every cued trial without confounding source memory and response bias. For every cued trial, participants saw the face studied with the test word as well as one of the faces in the opposite gender. The test instructions (correctly) informed them that one of the faces that they saw would always be the actual face studied with the test candidate. Thus, participants had no reason to disregard the external cues, which could perhaps result in a larger cueing effect than was observed in Experiment 3.

As in Experiment 3, participants studied words either with one male and one female face or with eight male and eight female faces. Participants should be easily able to internally reinstate the male and female face in the single-face condition; therefore, Experiment 4 should replicate Experiment 3 by showing no cueing effect in this condition. The multiple-face condition should create internal reinstatement difficulties; therefore, the cue discrepancy hypothesis predicts that external cues will improve source memory in this condition.

Experiment 4 Method

Participants. Data were collected from 92 Louisiana State University undergraduates who received course extra credit for volunteering. Participants were randomly assigned to the single-face and multiple-face conditions in equal numbers. With $f^2 = .02$ and $\rho = .4$, the power to detect an interaction effect was .69.

Design. This experiment conformed to a 2 (face number) \times 2 (cueing condition) design. Face number was manipulated between subjects with “multiple face” and “single face” as the individual levels. Cueing condition was manipulated within subjects, and the levels of this variable were “cued” and “uncued.”

Materials. This experiment involved the same words and pictures used in Experiment 3. Additionally, the study list had the same structure as in Experiment 3. The test list contained all 96 studied words, with half of the words studied in each gender appearing with no face at test and half appearing with face cues at test. On cued trials, the test word appeared near the top of the computer screen. Immediately below the word, a picture of a male face and a picture of a female face appeared side-by-side, with the female face on the left and the male face on the right. The face that matched the gender studied with the test item was always the individual face paired with the test item in the study phase. For the face that mismatched the gender studied with the test item, an individual face was selected at random with the constraint that all faces appeared an equal number of times.

Procedure. Procedural details matched Experiment 3 up to the instructions for the memory test. As in Experiment 3, test instructions informed participants that they would see words from the study phase and would rate their confidence that each word was studied with a male or a female face. Participants were told that they would see a male and a female face on

some of the test trials, and that one of the faces would always be the face studied with the test word. Participants in the multiple-face condition were told that any word appearing without faces at test could have been studied with any of the eight male or eight female faces from the study phase.

Experiment 4 Results and Discussion

Table 11 shows the hit rates, false alarm rates, and d' scores for both face number groups with and without retrieval cues. For the multiple-face group, providing cues resulted in a higher hit rate and a lower false alarm rate, and the d' scores reveal that participants were better able to discriminate male and female items on cued versus uncued trials. For the single-face group, retrieval cues had little or no effect on hit rate, false alarm rate, and d' . A 2 (face number) \times 2 (cueing condition) ANOVA produced a main effect of face number, $F(1, 90) = 4.87, p < .05$. As in Experiment 3, source performance was lower in the multiple-face condition ($M = .93$) than in the single-face condition ($M = 1.28$). The ANOVA also revealed a main effect of cue, $F(1, 90) = 5.31, p < .05$, and more critically, an significant interaction, $F(1, 90) = 7.99, p < .01$. Specific contrasts evaluating the cueing effect within each face number group showed that cues significantly affected source memory in the multiple-face condition, $t(45) = 3.57, p < .01$, but not in the single-face condition, $t(45) = .38, ns$. Thus, the results showed strong evidence of the interaction predicted by the cue discrepancy hypothesis.

The parameters of the signal detection model for Experiment 4 appear in Table 12. Mirroring the d' results, A_z scores were higher for cued versus uncued trials for the multiple-face group but unaffected by cueing condition for the single-face group. Equating memory parameters in the cued and uncued conditions resulted in a significant change in model fit for the

multiple-face group, $G^2(2) = 17.24, p < .001$. For the single-face group, assuming equivalent source memory across cueing conditions did not significantly impact model fit, $G^2(2) = .37, ns$.

TABLE 11

Observed recognition memory hit rates, false alarm rates, and d' scores across the face number and cueing conditions of Experiment 4

Cueing Condition	Face Number and Source Measure					
	16-Face			2-Face		
	HR	FAR	d'	HR	FAR	d'
Uncued	.66 (.02)	.38 (.02)	.76 (.10)	.73 (.02)	.28 (.03)	1.29 (.15)
Cued	.71 (.02)	.31 (.02)	1.11 (.11)	.73 (.02)	.29 (.02)	1.26 (.13)

Note: HR = hit rate (the proportion of male items called “male”); FAR = false alarm rate (the proportion of male items called “female”). Standard errors are in parentheses.

TABLE 12

Memory parameters produced by the signal detection models for the multiple- and single-face groups in Experiment 4

Cueing Condition	Face Number and Memory Parameter					
	16 Faces			2 Faces		
	μ	σ	Az	μ	σ	Az
Uncued	.94	.99	.75	1.33	1.07	.82
Cued	1.19	.95	.81	1.29	1.03	.82

Experiment 4 replicated the results of the third experiment in that retrieval cues improved source memory in the multiple-face condition but had no effect on source memory in the single-face condition. Moreover, the cueing effect was more pronounced in Experiment 4, and the interaction of face number and cueing condition reached significance. Removing trials in which only incorrect cues were provided increased the effectiveness of the cueing manipulation, allowing for more conclusive support for the cue discrepancy hypothesis.

GENERAL DISCUSSION

The experiments reported herein explored the interaction of external and internal reinstatement in source and recognition memory. Previous source and recognition experiments have found inconsistent effects of external context cues, with some studies providing evidence that cues benefit memory performance (Craik & Kirsner, 1974; Dodson and Shimamura, 2000; Geiselman & Bjork, 1980; Geiselman & Glenny, 1977; Kirsner, 1973; Murnane et al., 1999; Palmeri et al., 1993; Smith, 1986) and others demonstrating null effects of external cueing manipulations (Craik & Kirsner, 1974; Fernandez and Glenberg, 1985; Godden & Baddeley, 1980; Kirsner, 1973; Murnane & Phelps, 1993, 1994, 1995; Smith et al., 1978; Starns & Hicks, 2004; Starns & Hicks, 2005). According to the cue discrepancy hypothesis (Bjork & Richardson-Klavehn, 1989; Murnane & Phelps, 1994; Smith, 1979), this variability in results could reflect variations in the information available through internal reinstatement. External cues should improve memory performance only when they introduce information that is not available via internal reinstatement.

For the first two experiments, I developed MINERVA 2 (Hintzman, 1988) models to derive specific predictions regarding the interaction of external and internal reinstatement. The model results indicated that decreasing the feature overlap within a set of faces disrupted internal reinstatement. This result is best understood in terms of the source cue constructed by the model when no specific external cue is provided. Regardless of whether the model constructed an “average” cue summarizing the features of all faces within a set or simply reinstated a randomly-selected individual face, the face cue was a better match to the faces from the study phase with high feature-overlap than with low feature-overlap (see Figure 2, page 25). The differences in the effectiveness of internal reinstatement led to an interesting interaction with the external

cueing manipulation: external cues substantially improved memory performance with low feature overlap but had much smaller effects with high feature overlap. Thus, the model results were consistent with the cue discrepancy hypothesis; that is, robust cueing effects only emerged when internal reinstatement was disrupted by decreasing the feature overlap within a set of faces.

The empirical results did not support the model predictions. In both Experiments 1 and 2, external cues had no impact on memory performance for either the low- or high-overlap groups. Moreover, the results provided no evidence that the effectiveness of internal reinstatement was influenced by the overlap variable. Even on uncued trials, participants in the low-overlap group performed as well as participants in the high-overlap group. These results refute the formal models that I developed and fail to support the cue discrepancy hypothesis.

Experiments 3 and 4 were designed to provide more powerful tests of the cue discrepancy hypothesis than the first two experiments. Experiments 1 and 2 had no successful manipulation of internal reinstatement. In Experiments 3 and 4, I manipulated internal reinstatement by presenting either one or eight faces within each gender. The results supported the conclusion that increasing the number of faces impaired internal reinstatement: with no external cues, the multiple-face participants showed substantially lower source discrimination compared to the single-face participants. The effects of external cues matched the predictions of the cue discrepancy hypothesis. In both Experiments 3 and 4, external cues improved source discrimination for the multiple-face participants but not for the single-face participants. This pattern was most evident in Experiment 4, in which an improved cueing procedure enhanced the impact of external cues.

Overall, the results of this study support the cue discrepancy hypothesis; nevertheless, results were inconsistent with the more specific MINERVA 2 models. Why did the formal

models fail to produce accurate predictions? The models may have failed due to a lack of correspondence between the features that the models used to represent faces and the features of the actual stimuli. Other than the fact that the faces within a set shared more features in the high-overlap than in the low-overlap condition, the face vectors in the models had no direct relationship to the physical attributes of the faces seen by the participants. Moreover, face recognition studies suggest that people process faces primarily in terms of relations among features as opposed to the individual features themselves (Farah, Wilson, Drain, & Tanaka, 1998). Thus, the representational structure of MINERVA 2 may be inappropriate for face stimuli, which could explain why the models did not generate accurate predictions regarding the feature-overlap manipulation.

The models also may have failed to make successful predictions because their retrieval processes failed to capture participants' true internal reinstatement abilities. Specifically, the models assumed that, without external cues, participants could only reinstate a single face cue for each gender (i.e., the average face or a randomly selected individual face). Participants may have actually internally reinstated a number of individual faces on uncued trials, which could explain why uncued performance rivaled cued performance in Experiments 1 and 2. Indeed, if participants were able to internally reinstate all four faces in each gender, then external face cues would be completely unnecessary. Apparently, the eight faces/gender in Experiments 3 and 4 were enough to prevent the internal reinstatement of all individual faces on uncued trials, as evinced by the poorer uncued performance in the multiple-face versus the single-face conditions.

Another possible limitation of the MINERVA 2 models is that they represent a "single-process" approach to memory performance; that is, memory decisions in MINERVA 2 are based on a single continuous value summarizing all of the available memory evidence. In contrast,

many recognition theorists argue that memory performance reflects two underlying processes yielding distinct types of information. Specifically, recognition decisions can be based both on familiarity, a quantitative form of evidence resembling the evidence used in global matching models, and recollection, the recovery of specific qualitative details from the target event (Yonelinas, 1999). Notably, Macken (2002) has claimed that context reinstatement influences recollection but has no effect on an item's familiarity. The simulation results from the first two experiments show that a familiarity-based model can produce context effects, but such a model would be obviously inappropriate if reinstatement manipulations truly influenced only recollection. Although MINERVA 2's "single-process" nature somewhat limits the explanatory power of the model, this limitation is not the basis for the model's failed predictions in Experiments 1 and 2. Indeed, the results of Experiment 2 are as inconsistent with the dual process idea that retrieval cues affect only recollection as they are with the MINERVA 2 predictions. In Experiment 2, retrieval cues had no effect on overall recognition performance; thus, the cues had no impact on either familiarity *or* recollection. Adopting a dual-process perspective does not help to explain the null cueing effects.

Consistency with the Results of Existing Studies

Experiments 3 and 4 suggest that external cues influence source memory only in conditions that impair internal reinstatement. Some extant source memory studies appear to contradict this conclusion. For example, Dodson and Shimamura (2000) apparently demonstrated robust effects of source cues in a situation in which internal reinstatement should have been very effective. In a number of experiments, these researchers presented words in either a single male or a single female voice. At test, they compared conditions in which either no voice or an unstudied voice was externally reinstated at test to conditions in which a voice

from the study phase was provided as a cue. With only one voice per gender, participants should have been able to internally reinstate the voices from the study phase even when studied voices were not externally reinstated. Therefore, the paradigm is similar to the single-face condition from Experiment 3, and the cue discrepancy hypothesis predicts null cueing effects.

Consistent with the single-face results from Experiment 3, these researchers found that cueing with the studied male voice increased “male” responses to both male and female items and cueing with the studied female voice increased “female” responses to both male and female items. Thus, Dodson and Shimamura (2000) also found that gender cues impacted participants’ bias to use certain source responses. The results of Dodson and Shimamura depart from the results of Experiment 3 in terms of the model-based source memory measures designed to correct for these bias effects. I corrected for bias based on the equal-variance signal detection model (i.e., I computed d' scores), and the results revealed no effects of cues on source memory. Dodson and Shimamura used a multinomial model (Batchelder & Riefer, 1990) to correct for bias, and the model results showed that matching cues enhanced source memory. The demonstration of cueing effects with only one source exemplar in each gender seems to contradict the results of the single-face condition from Experiment 3 as well as the cue discrepancy hypothesis. However, the conclusion that cues affected source memory is only as valid as the model used to separate bias and memory effects, and a number of empirical findings have recently challenged the appropriateness of multinomial models of source memory (DeCarlo, 2003; Glanzer, Hilford, & Kim, 2004; Hilford, Glanzer, Kim, & DeCarlo, 2002; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick & Dodson, 2005; Slotnick, Klein, Dodson, & Shimamura, 2000).

The root of the problem with multinomial models is that they are based on the threshold retrieval of source information; that is, participants are assumed to identify in an absolute sense that a test candidate was studied in a certain source (Batchelder & Riefer, 1990). Failing source identification, participants are assumed to guess as to the source of the test candidate.

Considerable evidence is mounting that source retrieval is actually a continuous process (see review in Hilford et al., 2002), suggesting that multinomial models are inappropriate models of source memory. Therefore, Dodson and Shimamura's (2000) model results should be interpreted with caution. The results may only show an effect of retrieval cues on response bias that the multinomial model inappropriately attributed to memory differences.

A study by Palmeri et al. (1993) also appears to contradict the results of my Experiments 3 and 4 by showing that the number of source exemplars within a source category (such as the number of faces within the "female" category) does not interact with cueing effects. These researchers presented words in male and female voices in a continuous recognition paradigm. For each word that appeared, participants had to decide whether the word occurred previously in the sequence. In the second experiment, participants were also required to decide if repeated words were presented in the same voice or in a different voice as their prior presentation. Different groups of participants encountered 2, 6, 12, or 20 different voices in the experiment. The results showed a benefit in source accuracy when voices matched from a word's first to its second presentation; that is, participants were better able to identify matching voice repetitions as the same voice than they were to identify mismatching voice repetitions as a different voice. Moreover, the matching-voice advantage did not vary systematically with the number of voices encountered in the study phase. Palmeri et al. interpreted their results as demonstrating that source memory improved when voice cues matched a word's prior presentation, and also

concluded that the cueing effect did not interact with the total number of voices. This conclusion contrasts my conclusion that cueing effects emerge only when multiple exemplars are encountered within each source category.

Critically, as with Dodson and Shimamura (2000), the apparent cueing effects observed by Palmeri et al. (1993) may actually reflect only response bias. Participants were asked to discriminate matching from mismatching trials with “same” or “different” responses. A reasonable strategy given this test requirement would be to reinstate the voice heard upon the repeated presentation. Simultaneously cueing memory with the voice and item information on each trial would yield the amount of evidence that the current word was previously heard in the current voice. Participants could respond “different” if this evidence was low or “same” if this evidence was high. When framed in these terms, it becomes clear that participants would identify more matching than mismatching repetitions if they had a general bias to respond “same” rather than “different.” A memory-based and a bias-based account of the match/mismatch differences cannot be discriminated by the data reported by Palmeri et al., so the data reported by these researchers do not necessarily contradict my Experiments 3 and 4.

Future research is needed to determine if the effects I observed with faces generalize to other source characteristics, such as voice or location. The cue discrepancy hypothesis is not limited to a specific type of mnemonic information and presumably could be supported by experiments using a wide variety of sources. However, faces do have some special characteristics that should be considered in comparing faces to other source alternatives. Faces are complex, familiar, and meaningful stimuli, and as a result, participants may be able to easily integrate face and word information. Indeed, the study instructions of the present experiments explicitly encouraged participants to integrate face and word information (i.e., they were told to

imagine the face saying the word or acting out the word). A high degree of integration could make memory for faces more acute than memory for some other types of sources, and could also make faces more powerful cues (see Murnane et al., 1999). Regardless, any characteristic for which participants display accurate source memory must have some degree of integration with word information; therefore, reinstating the characteristic should improve source memory as long as the information cannot be effectively internally reinstated. Although cueing effects may not be as large for alternative types of source characteristics, the effects should not be absent.

If Dodson and Shimamura's (2000) results indeed show that cues affect source memory, the critical difference between this study and my single-face conditions may be the use of voices instead of faces as source exemplars. Participants may be less capable of reinstating voices without external support because they cannot vividly remember the voices that they heard in the study phase. In contrast, participants may have very detailed memory for faces seen in the study phase, and this detailed memory may translate to very accurate internal cues. If voice memory is less detailed than face memory, then internal reinstatement of voices may be ineffective even with only one voice in each gender, which would explain why voice cues (purportedly) improved source memory in Dodson and Shimamura.

Global Matching Models and Episodic Memory

The MINERVA 2 models developed for Experiments 1 and 2 produced results that contrasted with the observed data; therefore, the formal modeling in this study was not particularly enlightening concerning the interacting effects of feature overlap and cueing. However, global matching models have proved to be valuable tools for exploring context memory in other studies (Dennis & Humphreys, 2001; Dodhia & Metcalfe, 1999; Hintzman, 1988; Metcalfe, 1990; Ratcliff et al., 1995), including studies conducted by the current author

(e.g., Hicks & Starns, in press). This contrast suggests that some aspects of global matching models are more useful than others. For example, global matching models represent stimuli with simple featural structures that rarely have any formal correspondence to the characteristics of actual stimuli. Moreover, encoding simply consists of probabilistically copying these featural structures into memory traces. For these reasons, global matching models may fail to provide reasonable explanations for the effects of encoding variables in memory experiments – the models simply do not have the mechanisms necessary to capture the complexity of encoding procedures.

In contrast, the retrieval mechanisms of global matching models may be more successful in providing novel explanations and stimulating incisive experiments in the memory literature. In the models, memory assessments are directed to specific target events based on the information that is reinstated in the retrieval probe. For example, to assess the memory evidence that a word was encountered with a male face, the features of the word and the features of a male face are reinstated at retrieval. Thus, the models suggest that episodic memory performance is limited by the ability to form context-specific memory probes. In Experiments 3 and 4, presenting multiple faces within each gender apparently impaired participants ability to form effective source probes without external cues, and future research can explore other variables that disrupt internal source cueing. Another enlightening property of global matching models is the manner in which they express mnemonic evidence. The retrieval mechanisms of global matching models summarize memory evidence into a single quantitative variable (such as echo intensity in MINERVA 2). This characteristic of the models provides a compelling psychological interpretation for mathematical models that analyze source performance in terms of a continuous evidence variable (such as signal detection theory). In short, global matching

models probably cannot provide a complete account of source memory performance; nevertheless, certain properties of the global matching perspective can continue to guide theoretical insights into source memory phenomena.

Conclusion

Many otherwise diverse theories of memory share a common emphasis on the importance of retrieval cues (see Roediger, 2000). In general, memory theorists agree that event memory should be most effective when the cues available at retrieval closely correspond to the characteristics of the target event (e.g., Tulving & Thomson, 1973). Nevertheless, a number of empirical results demonstrate null effects of experimenter-provided cues on memory performance (Craik & Kirsner, 1974; Fernandez and Glenberg, 1985; Godden & Baddeley, 1980; Kirsner, 1973; Murnane & Phelps, 1993, 1994, 1995; Smith et al., 1978; Starns & Hicks, 2004; Starns & Hicks, 2005). The current study takes a step toward resolving this conflict between theory and data in terms of the cue discrepancy hypothesis, which states that the success of cueing manipulations depends on the cues that are internally reinstated when external cues are unavailable. The results of Experiments 3 and 4 support the contention that retrieval cues *always* powerfully impact source memory performance, but that participants are not limited to the external cues manipulated by the experimenter. When participants can easily remember the specific faces that were potentially studied with a word (i.e., when they study only one male and one female face), they can internally reinstate these faces as part of their source decisions. Thus, effective face cues are available regardless of the current level of the explicit cueing manipulation. When participants encounter ambiguity as to which specific face within a gender may have been studied with a word (i.e., when they study eight male and eight female faces), they cannot internally reinstate highly effective face cues. As a result, they become more

dependent on the external cues, and significant cueing effects emerge. Ironically, context reinstatement manipulations can fail because effective context cues are so critical to memory performance that participants internally reinstate context information even when they are not explicitly prompted to do so. To successfully investigate cueing effects, researchers must carefully design their experiments to ensure that the information in external cues is not available via internal reinstatement.

REFERENCES

- Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.
- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In Chizuko Izawa (Ed.), *Current Issues in Cognitive Processes: The Tulane Flowerree Symposium on Cognition* (pp. 313-344). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Buchner, A., Erdfelder, E., & Faul, F. (1997). *How to use G-Power*. http://www.psych.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274-284.
- DeCarlo, L. T. (2003). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 767-778.
- Dennis, S. & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- Dodhia, R. M. & Metcalfe, J. (1999). False memories and source monitoring. *Cognitive Neuropsychology*, *16*, 489-508.
- Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1023-1044.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1-11.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, *105*, 482-498.
- Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, *13*, 333-345.

- Geiselman, R. E., & Bjork, R. A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, *12*, 188-205.
- Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, *5*, 499-504.
- Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1176-1195.
- Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99-104.
- Hicks, J. L., & Starns, J. J. (in press). Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General*, *131*, 494-510.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411-428.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.
- Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*, 208-233.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source Monitoring. *Psychological Bulletin*, *114*, 3-28.
- Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition*, *1*, 449-453.
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and experimental contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297-315.
- Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 153-161.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A users guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning & Verbal Behavior*, 8, 410-423.
- Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., & Neuberg, S. L. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology*, 85, 1107-1120.
- Metcalfe, J. (1990). Composite Holographic Associative Recall Model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General*, 119, 145-160.
- Murnane, K. & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 882-894.
- Murnane, K. & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, 22, 584-590.
- Murnane, K. & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 158-172.
- Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, 128, 403-415.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309-328.
- Psychological Software Tools (2002). E-prime. <http://www.pst-net.com>.
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1110-1115.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: General*, 124, 353-374.
- Roediger, H. L. III. (2000). Why retrieval is the key process in understanding human memory. In Endel Tulving (Ed.), *Memory, Consciousness, and the Brain: The Tallinn Conference* (pp. 52-75). New York, NY: Psychology Press.
- Slotnick, S. D. & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, 33, 151-170.

- Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1499-1517.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460-471.
- Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition*, *14*, 347-354.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342-353.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203-220.
- Starns, J. J., & Hicks, J. L. (2004). [Context reinstatement effects on source memory with encoding tasks that encourage item/context integration]. Unpublished raw data.
- Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional source-monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1213-1220.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, *62*, 74-82.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415-1434.

VITA

Jeffrey Starns was raised in Hammond, Louisiana, and he graduated from Loranger High School in 1997. He received a Board of Trustees scholarship to attend Southeastern Louisiana University, and he graduated with a Bachelor of Arts degree in psychology in 2001. Upon graduation, he received the Psychology Award for having the highest GPA of all graduating psychology majors, and he also received the President's Award for Academic Excellence for being the top graduating student in the College of Arts and Sciences. He earned his master's degree in psychology from Louisiana State University. He was awarded the LSU Dissertation Fellowship for the 2005-2006 academic year. He has published several articles on human memory in professional journals, and aspires to be a professor at a research-oriented university.