Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2010 January ; 36(1): 66–79. doi:10.1037/a0017394.

The Role of Memory Activation in Creating False Memories of Encoding Context

Jason Arndt

Department of Psychology, Middlebury College

Abstract

Three experiments examined false memory for encoding context by presenting DRM themes (Deese, 1959; Roediger & McDermott, 1995) in usual-looking fonts and testing related, but unstudied, lure items in a font that was shown during encoding. In two of the experiments, testing lure items in the font used to study their associated themes increased false recognition relative to testing lure items in a font that was used to study a different lure's theme. Further, studying a larger number of associates exacerbated the influence of testing lure items in a font used to study their associated themes. Finally, testing lures in a font that was encoded many times, but was not used to present the lures' studied associates, increased lure errors more than testing lures in a font that was encoded relatively fewer times. These results favor the explanation of false recognition offered by global-matching models of recognition memory over the explanations of activation-monitoring theory and fuzzy-trace theory.

Keywords

false recognition; encoding context; models of recognition memory

Human memory is subject to a multitude of errors, including source misattributions (Zaragoza & Lane, 1994), distortions (Schacter, 1995), and the creation of false memories (Loftus, 1993; Roediger & McDermott, 1995). One of the most useful bases for both aiding accurate memory and distinguishing authentic memories from false memories is encoding context (Geiselman, Fisher, Cohen, Holland, & Surtes, 1986; Lindsay & Johnson, 1989; Smith, 1979). For example, the cognitive interview (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avettisan, & Prosk, 1984) utilizes context reinstatement to aid witnesses' memory for real-world events. Importantly, the cognitive interview also reduces the negative effects of misinformation, presumably in part due to context reinstatement (Geiselman, et al., 1986). Similarly, techniques that require participants to judge the source of their memories reduce the negative impact of post-event misinformation (Lindsay & Johnson, 1989). Thus, retrieval of encoding context can not only improve how well authentic memories are remembered, but can also reduce memory errors.

The research presented in this paper explores the possibility that reinstating encoding context at retrieval may not always reduce memory errors, and in fact, may exacerbate some types of

Correspondence should be addressed to Jason Arndt, Department of Psychology, 5605 Middlebury College, Middlebury, VT, 05753. jarndt@middlebury.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/xlm.

errors. This issue is explored using the Deese-Roediger-McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm. Memory errors in the DRM paradigm are caused by presenting participants with a series of words (referred to as *themes* hereafter) that are all related to a single unstudied word, known as the *lure item*. For example, participants may study words such as *chirp*, *sparrow*, *bluejay*, *canary*, *feathers*, *nest*, *pigeon*, and *robin*. On a later recognition memory test, participants are shown some of the words they actually studied, as well as *bird*, the lure item. Although most recognition memory errors are familiarity-based (e.g., Jones & Jacoby, 2001), lure errors show several characteristics that suggest participants believe they can recollect lures' occurrence on a study list. First, lure error rates are much higher than other recognition memory errors, approaching the level of hits under some conditions (Roediger & McDermott, 1995). Second, participants report that lure errors are often accompanied by recollection phenomenology (Roediger & McDermott, 1995; Tulving, 1985). Third, participants are willing to attribute lures to a specific source that was encountered during encoding (Hicks & Hancock, 2002; Hicks & Starns, 2006a; Roediger, McDermott, Pisoni, & Gallo, 2004), suggesting they believe that they retrieved episodic details of the lures' presentation. Fourth, participants make *more* errors to lure items when they are given a source memory task than when they are simply given an old-new recognition task (Hicks & Marsh, 2001). Thus, when participants are required to scrutinize their memories for specific details of study item presentation in order to make a source memory judgment, they are actually more error-prone than when they are not required to do so. These regularities suggest that the results of memory search for lures leads participants to believe lures occurred in the same encoding context as studied items. Specifically, if participants find evidence in their memory search that lure items were encountered during encoding, they would be expected to show high rates of lure errors (Roediger & McDermott, 1995), to believe they can recollect lures' presentation on a study list (Roediger & McDermott, 1995), and to be willing to attribute lures to a specific source that was encountered at study (Hicks & Hancock, 2002; Hicks & Marsh, 2001; Roediger, et al., 2004). If it is true that testing memory with lure items inspires retrieval of encoding context, testing lure items in contexts that were experienced during encoding should exacerbate lure errors. This general question served as the empirical impetus for the studies that are reported in this paper.

A second motivation for the studies in this paper was to test three theoretical frameworks that have been used to explain lure errors in the DRM paradigm: Activation-monitoring theory (Roediger, Watson, McDermott, & Gallo, 2001), fuzzy-trace theory (Brainerd, Reyna, & Kneer, 1995), and global matching models (Hintzman, 1988; Shiffrin & Steyvers, 1997). Each of these theories explains why people make lure errors, as well as why people behave as if lures were experienced in the same encoding context as studied items, but do so using different underlying assumptions. Thus, distinguishing between their explanations of memory errors can aid our understanding of the underlying bases of associative memory errors, as well as why people find lures in the DRM paradigm subjectively compelling.

Activation-monitoring theory suggests that studying a lure's associates activates the lure's representation in semantic memory, which in turn increases the probability that participants will make errors to that lure on a subsequent memory test (Roediger, et al., 2001). Although activation is the primary process in activation-monitoring theory that produces lure errors, activation alone is insufficient to explain why people behave as if lures were experienced in a particular encoding context. For example, consider the fact that people attribute lures to the source in which their studied associates were experienced (Roediger, et al., 2004) and that people tend to believe that lures were presented in a source that was correlated with the lure's strongest associates (Hicks & Hancock, 2002; Hicks & Starns, 2006a). The activation level of a lure's representation does not contain information that specifies the source of its activation, rendering activation alone an insufficient basis for making systematic source judgments to lure items. Thus, the ability to explain lure source judgments requires a mechanism by which

participants can not only assess the activation of a lure's representation, but also the source that was associated with that activation during encoding. The most straightforward mechanism that allows activation-monitoring theory to account for such a result is that when a lure's representation is activated strongly during encoding, it can become associated with encoding context (Roediger, et al., 2001; 2004). Thus, this mechanism leads activation-monitoring theory to suggest that to the extent participants believe lure items are associated with encoding context, it may be because lures' representations in memory *are* associated with encoding context.

Fuzzy-trace theory (Brainerd, et al., 1995) proposes that memory traces are stored on a continuum that ranges from verbatim information about the experience to gist information about the experience. Verbatim traces represent the perceptual details of an experience, such as visual features, which allow people to differentiate memories from one another. Thus, retrieval of verbatim traces underlies the ability to recollect study item occurrence in fuzzy-trace theory. Gist information, on the other hand, represents the commonalities among experiences that are part of an episode, and underlies test item familiarity. As applied to the DRM paradigm, gist information represents the semantic commonalities among a lure's studied associates, which leads fuzzy-trace theory to propose that lure errors are familiarity-based. In general, fuzzy-trace theory suggests that memory errors to unstudied items arise from how well they match gist traces, and that memory errors are limited by the extent to which unstudied items produce retrieval of verbatim traces. Thus, lure errors increase when they match the gist representation of their studied associates and decrease when they inspire retrieval of the verbatim traces of their studied associates. Finally, fuzzy-trace theory proposes that if a gist trace that is retrieved is sufficiently strong, it can produce a phenomenon known as *phantom recollection* (Brainerd, Wright, Reyna, & Mojardin, 2001). When phantom recollection occurs, participants confuse the strength of a gist trace with the psychological experience of recollecting, which is normally mediated by retrieving verbatim traces of studied items. Thus, according to fuzzy-trace theory, people believe they can recollect DRM lures because lure items tend to match very strong gist traces in memory. In essence, fuzzy-trace theory proposes that although lure errors are often phenomenologically similar to items that were episodically experienced, lures' recollection phenomenology is representationally distinct from that of study items.

Finally, global-matching models propose that lure errors are not the result of retrieving a specific memory representation (e.g., their activated representation in a semantic network or the gist trace of an experience), but instead occur because lures are similar to numerous traces in memory – those of their encoded associates. For purposes of simplicity, MINERVA2 (Hintzman, 1988) is used to illustrate the account offered by global-matching models, because it has been used to explain lure errors in the DRM paradigm (Arndt & Hirshman, 1998; Hicks & Starns, 2006b). Encoding in MINERVA2 involves the storage of memory traces for each study item. At retrieval, test items are compared to all of the traces in memory. Comparing a test item to each item in memory generates an activation value based upon the similarity between the test item and each memory trace. The activation values created by this comparison are then summed across all traces in memory, resulting in a value representing the memory activation that is caused by probing memory with the test item. If memory activation is sufficiently high, the test item is judged old. MINERVA2 suggests that study item memory arises primarily from the match of a test item to its encoded trace in memory. However, lure errors arise from summing the series of small matches that occur because the lure has a small amount of similarity to the memory traces of its studied associates (Arndt & Hirshman, 1998; Hintzman, 1988). Thus, even though lure items may not strongly match any individual trace in memory, summing the many smaller matches a lure item has with traces of its associates in memory can result in substantial activation of memory as a whole, producing compelling evidence that a lure was studied.

Global-matching models explain memory for encoding context in a similar manner to how they explain recognition memory in general. These models suggest that memory representations are composed of both item information (e.g., semantics) as well as contextual information (e.g., voice of presentation; Hicks & Starns, 2006b). At retrieval, a memory probe is constructed that includes both the item information in the test cue and contextual information present at test, which is then compared to all of the item and context traces in memory. Thus, the extent to which the test probes match item information and contextual information in memory provides an index of how likely it is the test item was studied in a particular context. Further, global-matching models possess a characteristic known as *interactive cueing* (Clark & Gronlund, 1996; Hicks & Starns, 2006), which causes memory probes that match both item and context information in a single trace to generate a larger activation signal than memory probes that activate item information and context information in different memory traces. Interactive cueing allows global matching models to explain memory for encoding context because a test item that is retrieved in the same context in which it was encoded will match both item and context information in the same memory trace, while a test item that is retrieved in a context different from the one in which it was encoded will match item and context information in different memory traces, producing more memory activation in the former case than in the latter. Similar to their explanation of lure errors in general, global-matching models propose that participants' belief that lure items were experienced in a given encoding context results from the summation of small matches of the lure to item traces and context traces in memory. Critically, because of interactive cueing, the matches of a lure to both the item and context portions of its associates' memory traces provides the basis for people to believe lure items were experienced in the same encoding context as their studied associates (Hicks & Starns, 2006b).

The Present Studies

In the present experiments, study items from DRM themes were encoded in specific visual contexts, unusual-looking fonts (Arndt, 2006; Arndt & Reder, 2003; see Figure 1 for examples). At test, study items and lures were presented in font contexts that matched the font context used to encode the lure's theme during study or in a font context that was encoded, but was used to encode a different theme during study (i.e., a mismatching font context). Prior research suggests that participants believe lures were presented in the font used to study their associates (Arndt, 2006). In particular, lure errors were higher when they were tested in a font that was used to study their associates compared to when they were tested in a font that was not presented during encoding (Arndt, 2006, Experiment 2). This result, like the four regularities cited above, suggests that participants believe lures were actually presented during encoding, because matching font of presentation from study to test enhances recognition memory and tends to have its effects on recollection phenomenology (Reder, Donavos, & Erickson, 2002).

As with their general explanation of false memory for encoding context, the three theories reviewed above explain this outcome, but do so using different assumptions. Fuzzy-trace theory suggests lure errors in the DRM paradigm occur because lure items match very strong gist traces. Although gist traces in the DRM paradigm generally represent the *semantic* commonalities that result from studying a number of related items (Brainerd & Reyna, 2005; Brainerd, et al., 2001; Reyna & Lloyd, 1997), it is also possible that fonts, as well as other sources commonly used in memory experiments, can produce gist representations of their semantic or perceptual commonalities when they are repeatedly presented during encoding (Brainerd & Reyna, 2005).¹ Thus, fuzzy-trace theory claims that lure errors are higher when lures are tested in a font used to study their associates because those memory probes will match both DRM themes' *semantic gist*, as well as the gist of the fonts presented at encoding (referred

¹I thank Chuck Brainerd for raising this point during the review process.

to as *font gist* hereafter for expositional clarity). In comparison, lures tested in an entirely unstudied font will only match semantic gist, producing lower error rates compared to the condition where lures were tested in a font used to study their associates.

Activation-monitoring theory can explain this result by suggesting that lure representations become associated with experimental context when they are activated during encoding. As a result, probing memory with a lure item and the font used to encode its associates will retrieve evidence that the lure was encoded in a particular font during the study list. Thus, when a lure's associates were studied in the font used to present the lure on a memory test, lure errors would be expected to be greater than when a lure was tested in a font not shown at encoding, because matching perceptual format between study and test improves recognition memory (Reder, et al., 2002).

Finally, global-matching models claim that encoding results in the storage of individual memory traces that represent both the items that were encountered, as well as the context in which they were encountered (e.g., the font used to present items at study). As a consequence, lure items match both item and context traces in memory when they are tested in a font that was experienced during encoding, producing more memory activation than when lures do not match context traces, as would occur when lures are tested in a font that was not experienced during encoding.

Although all three theories account for the finding that lure errors were higher when they were tested in a font used to show their studied associates, other evidence favors fuzzy-trace theory over activation-monitoring theory and global-matching models. Specifically, when lure items were tested in a font that was studied, but was not used to present the lure's associates during encoding, lure errors did not differ compared to a condition where lure items were tested in the font used to study that lure's associates. (Arndt, 2006; Experiment 1). Fuzzy-trace theory accounts for this result because it suggests that separate gist representations are stored for different aspects of experience, such that semantic gist represents the semantic commonalities of the words in each DRM theme, while font gist represents the commonalities of the fonts used to present study items. The argument for separate storage of semantic and font gist representations is that the *content* of the information that is represented in the two types of gist traces differs, which gives rise to separate representations in much the same way that different DRM themes give rise to different semantic gist representations and that a single item with multiple meanings can give rise to multiple gist representations (Brainerd & Reyna, 2005). Given the supposition of separate gist representations for the semantics and font of each theme, fuzzy-trace theory predicts similar levels of errors when a lure item is tested in a font used to study its associates compared to when a lure item is tested in a font that was studied, but was used to present a different lure's associates at encoding. This prediction occurs because lures that are tested in the font used to study their associates and lures that are tested in a font used to study a different lure's associates both match semantic gist arising from the encoding of the lures' associates and font gist arising from the encoding of a font during the study list. As a result, fuzzy-trace theory's explanation of DRM lure errors in terms of gist representations suggests that people will not be able to determine the *specific* encoding context that was used to study a lure's associates. Thus, although gist representations can underlie two different types of phenomenology, familiarity and phantom recollection, gist representations maintain one of the core properties of representations that mediate familiarity in dual-process models – that familiarity provides an acontextual measure of the likelihood that an event occurred (see Yonelinas, 2002 for a review). Retrieval of contextual information in fuzzy trace theory, like other dual-process models of recognition, requires access to representations that mediate *authentic* recollection, which in the case of fuzzy-trace theory are verbatim memory traces.

In contrast to fuzzy-trace theory, both activation-monitoring theory and global-matching models predict that testing a lure item in a font used to study its associates should increase lure errors compared to a condition where lures were tested in a font that was shown during study, but was not used to present words related to the lure. Activation-monitoring theory makes this prediction because lure representations are active, and thus prone to become associated with encoding context, while their associates are being encoded. As a consequence, the theory predicts that lure representations should be *specifically* associated with the font in which their associates were studied, and not other fonts experienced during encoding. Thus, when a lure is tested in the font used to study its associates, that lure representation's association with the font used to study its associates will be cued. In comparison, when a lure is tested in a font used to study a different lure's associates, the font-lure association will not be cued, and only lure representation activation will mediate lure errors.

Global matching models make the same prediction as activation-monitoring theory, but for very different reasons. In global-matching models, this prediction is a consequence of interactive cueing. When lure items are tested in the font used to study their associates, they will match their associates' item memory traces as well as the font (context) representation of their associates, producing interactive cueing. In comparison, when lure items are tested in a font that was used to study another lure's associates, the lure's memory probe will match the item traces of its associates and the font (context) representation of memory traces for a different lure's associates. Because the item and font information the memory probe activates is contained in different memory traces, interactive cueing will not occur and the overall activation of memory will be less than when lures are tested in the font in which their associates were studied. Thus, global-matching models, like activation-monitoring theory, argue that memory search produces evidence that lures were encountered in a *specific* encoding context, although the basis for that belief, how well lures match the traces of their encoded associates, differs from the mechanism posited by activation-monitoring theory.

Given that the three theories make different claims about the specificity with which lure representations are associated with encoding context, and different assumptions underlie those claims, further examining the extent to which lure errors are associated with encoding context can aid understanding of the bases of memory errors as well as why people find DRM memory errors subjectively compelling. To this end, the studies reported in this paper tested the specificity with which lures are associated with the encoding context of their studied associates by manipulating whether lures were tested in a font used to study their associates or in a font that was studied, but was used to study a different lure's associates. In addition, the studies reported in this paper examined whether font match effects for lures interacted with a variable that reliably increases lure errors, the number of associates studied (Arndt & Hirshman, 1998; Robinson & Roediger, 1997). Importantly, studying how font-match effects interact or do not interact with the number of associates studied rigorously tests the accounts of DRM errors advanced by fuzzy-trace theory, activation-monitoring theory, and global matching models. The specific tests of these theories provided by each experiment are detailed prior to each experiment.

Experiment 1

Two factors were manipulated in Experiment 1: The number of associates studied (two or eight), and whether the font used to present a theme's test items was used to present that theme's study items during encoding, or if the font used to present a theme's test items was studied, but was used to show a different theme's study items during encoding (referred to as *study-test match* hereafter). As reviewed above, both activation-monitoring theory and global-matching models predict that lure errors will increase when they are tested in the font used to study their associates, while fuzzy-trace theory expects lure errors will not be affected by this

manipulation. Experiment 1 also tested more detailed predictions of the three theories by manipulating the number of a lure's associates that were studied. Both activation-monitoring theory and global-matching models predict that lure errors should show a larger study-test match effect when more associates are studied. In activation-monitoring theory, this prediction arises because the theory proposes a relationship between lure activation and how strongly a lure will be associated with encoding context. Thus, the more a lure representation is activated during encoding, as would happen when more of its associates are studied, the stronger its association will be with the presentation context of its studied associates. Global-matching models also make this prediction, but do so because of their interactive cueing assumption. Specifically, the larger the number of item and context traces that a lure's memory probe activates, the greater the disparity there should be between a condition where a lure's memory probe matches item and context information that are part of the same memory trace (i.e., the match condition) compared with a condition where a lure's memory probe matches item and context information that are part of different memory traces (i.e., the mismatch condition). As an illustration of this prediction, assume that when a lure's memory probe matches item and context information that are part of the same memory trace, it produces an activation value of 1.5 (in arbitrary units of activation), but that when a lure's memory probe matches item and context information that are part of two different memory traces, it produces an activation value of 1.0. Thus, interactive cueing increases memory activation by 0.5. Further, assume that either two or eight of a lure's associates were studied during encoding. In this example, testing a lure in the font used to study its associates would produce a small amount of extra activation when two associates are studied (1.0 units), but a much larger amount of extra activation when eight associates are studied (4.0 units).

Finally, fuzzy-trace theory makes two predictions for this experiment. First, lure errors should be higher when more of a lure's associates are studied, because both the semantic gist trace and font gist trace that are formed when more associates are studied should be stronger. Second, study-test match should not impact lure errors because semantic and font gist traces will be accessed when lure items are tested in both the match and mismatch conditions.

Method

Participants

Participants in Experiment 1 were 64 Middlebury College students who participated in exchange for \$10 payment or as part of an Introduction to Psychology research appreciation requirement.

Materials and Design

Stimulus items were selected from Nelson, McEvoy, and Schreiber (1998). Sixty-four sets of eight words (themes) were chosen to serve as study items. Themes were chosen such that all of the items within a theme produced the same word (the lure) in free association with a nonzero probability (mean = 0.489; range: 0.055 to 0.960; range of mean associative strength for themes: 0.388 to 0.702). For example, the stimulus items *chirp*, *sparrow*, *bluejay*, *canary*, *feathers*, *nest*, *pigeon*, and *robin* were chosen as study items, and all produce the lure item *bird* in free association. In addition, 64 unrelated items were selected from Kucera and Francis (1967) to serve as new items on the recognition memory test. These items ranged in word frequency from 50 to 245 occurrences per million, with an average frequency of 99.53. Sixty-four unusual-looking fonts used in Arndt (2006) were selected to present study and test items. Fonts chosen for stimulus presentation were selected because they did not look strongly like other fonts used in this study, and because they were unlikely to have been experienced by participants prior to this experiment (Figure 1).

There were two manipulations in this experiment, *Number of Associates Studied* (two vs. eight) and *Study-Test Match* (match vs. mismatch). In order to implement these two manipulations, the sixty-four themes were divided into four sets of sixteen themes. The four sets of themes were rotated through the experimental conditions formed by crossing the Number of Associates Studied and Study-Test Font Match factors, such that each set served equally often in the four experimental conditions. For two of the sets of themes, all eight study items from the theme were presented as study items. For the other two sets of themes, two of the eight study items were randomly selected to be presented as study items. When a theme was assigned to the two associates studied condition, the two study items presented for that theme were randomly chosen anew for each participant. Half of each set of themes was combined to form each of two study lists, producing a study list length of 160 items ($16 \text{ themes} \times 8 \text{ study items} + 16 \text{ themes} \times 2 \text{ study items}$). All of the study items in each theme were presented in a single font, with a different font assigned to each of the sixty-four studied themes. Assignment of fonts to themes was determined randomly for each participant. Study items were presented blocked by theme, with the assignment of themes to serial positions within the study list and study items to serial positions within a theme determined randomly for each participant.

On the test list following each study list, two studied items from each theme, the lure item from each theme, and thirty-two unrelated test items were presented, producing a test list length of 128 words ($32 \text{ themes} \times 2 \text{ studied items per theme} + 32 \text{ lures} + 32 \text{ new items}$). When two associates from a theme were presented during the study list, those same two associates were presented on the recognition memory test. When eight associates from a theme were presented on a study list, two of those eight associates were randomly chosen to be presented on the recognition memory test. Further, the test items for half of the themes (both studied items and the theme's lure item) within each level of number of associates studied were tested in the same font that was used to present the theme's study items on the study list (the *match* condition), such that studied items were tested in the same font they were studied in, and the theme's lure item was tested in the font used to study its associates. The test items for the other half of the themes within each level of number of associates were tested in a font that was shown on the study list, but was not used to present the theme's study items on the study list (the *mismatch* condition). Thus, studied items were tested in a font that was studied, but was used to present a different theme during the study list, and lure items were tested in a font that was used to present a different theme's associates on the study list. Fonts were assigned to test items such that it was always the case that mismatching fonts came from the same level of Number of Associates Studied. Thus, test items assigned to the eight associates studied condition were tested in a font used to show eight associates during encoding, and test items assigned to the two associates studied condition were tested in a font used to show two associates during encoding. Finally, each new item was presented in one of the fonts used on its test list. Thus, new items provided a baseline assessment of false alarm rates when a font was presented eight times on a study list relative to when a font was presented twice.

Procedure

Participants completed two study-test lists, in which all aspects of the procedure were identical. Prior to the first study list, participants were instructed that they would see a series of words one at a time on a computer screen, and that the words would be shown in a variety of unusual-looking fonts. They were asked to rate the appropriateness of the font for the meaning of the word (Arndt, 2006; Arndt & Reder, 2003; Reder, et al, 2002) on a four point scale, where 1 indicated "not very appropriate" and 4 indicated "very appropriate." For example, if a study word was *breeze* and the font in which *breeze* was presented was italicized, indicating the movement in a flexible object a breeze might create, they would be likely to judge that font-word correspondence as "appropriate" or "very appropriate." Participants were further instructed that there were no right or wrong answers on the font-word judgment task, but that

their perceptions of font-word correspondence were of interest. Finally, participants were instructed to do their best to remember each word they judged, because there would be a memory test following presentation of the study list. Study list presentation then commenced with each study item being presented serially in the center of the computer screen. Study items remained on the screen until participants judged an item's font-word correspondence.

Upon completion of the first study list, participants were given an old-new recognition memory test for the words from the first study list. Participants were instructed to determine whether each word was shown on the preceding study list by pressing the “o” key for “old” or the “n” key for “new.” Test items were presented serially on the computer screen in a random order. Upon completing the first recognition memory test, participants were then shown a second study list, followed by a second recognition memory test. Prior to presentation of the second study list and second test list, they were instructed that all aspects of the procedure were identical to the first study and test list, and that they no longer needed to remember items from the first study list, because their memory for those words would not be tested again.

Results

The results of Experiment 1 were analyzed with 2×2 within-subjects ANOVAs for each of two dependent measures. First, the probability a test item was judged as studied was analyzed separately for studied items, lure items, and new items. Second, new item false alarms were used as a measure of baseline error rates for computing d' for both studied items and lure items. While it may seem unorthodox to compute a measure of discriminability for lure items, computation of d' provides a measure of how much lure errors exceeded the baseline false alarm rates to new items. Such computations will ensure patterns in lure errors are not due to differences in overall error rates across the font presentation frequencies (two vs. eight). Only analyses of the probability test items were judged “old” are presented for Experiment 1, because analyses of d' led to the same conclusions as those of probability “old” judgments. The alpha level for analyses of all data reported in this paper was .05.

New item false alarms were generally low, such that studying a font eight times ($M = .17$) did not increase new item false alarms reliably more than studying a font twice ($M = .15$; $t(63) = 1.91$, $p = .061$). Figure 2 presents the probability of judging study items “old” (top panel) and the probability of judging lure items “old” (bottom panel) as a function of Number of Associates Studied and Study-Test Match. For *studied* items, there was a main effect of Study-Test Match, $F(1,63) = 109.40$, $MSE = .005$, $p < .001$, such that items in the match condition ($M = .92$) were recognized better than items in the mismatch condition ($M = .84$). Neither the main effect of Number of Associates Studied, $F(1,63) = 3.32$, $MSE = .005$, $p = .073$, nor the interaction between Study-Test Match and Number of Associates Studied, $F(1,63) = 1.04$, $MSE = .005$, $p = .313$, reached the criterion for significance. For *lure* items, there was a main effect of both Study-Test Match, $F(1,63) = 38.17$, $MSE = .011$, $p < .001$, and Number of Associates Studied, $F(1,63) = 222.12$, $MSE = .022$, $p < .001$, as well as an interaction, $F(1,63) = 7.50$, $MSE = .011$, $p = .008$. The main effect of Study-Test Match demonstrates that lure errors were higher when they were tested in the font used to present their associates during encoding (i.e., the match condition; $M = .50$) than when they were tested in a font used to present a different lure's associates during encoding (i.e., the mismatch condition; $M = .43$). The main effect of Number of Associates Studied demonstrates that lure errors were higher when eight of a lure's associates were studied ($M = .60$) than when two of a lure's associates were studied ($M = .32$). Finally, the interaction indicates that testing a lure in the font used to study its associates had a larger effect in the eight associates studied condition than in the two associates studied condition (consult Figure 2).

Discussion

The results of this experiment are consistent with the explanation of DRM lure errors offered by activation-monitoring theory and global-matching models, but are inconsistent with the explanation of fuzzy-trace theory. Thus, these results favor theories of false memory that suggest lure items cue retrieval of specific information about the encoding context of their studied associates, as both activation-monitoring theory and global-matching models claim. However, there are two alternate explanations of the interaction between number of associates studied and study-test font match that can undermine the support the data provide for activation-monitoring theory and global matching models. First, the interaction could represent a scale effect (Loftus, 1978), because lure errors in the two associates condition were considerably lower than they were in the eight associates condition. Thus, there was less room for lure errors to be affected by study-test font match in the two associates condition, such that the interaction between number of associates studied and study-test font match may be an artifact of overall error rate differences between the two and eight associates conditions. Second, it is possible that more than only activation differences (lure activation or memory activation) differed across the two and eight associates conditions. Recent evidence suggests that people are better able to employ recall-to-reject processing in the two associates match condition than in the two associates mismatch condition or either of the eight associates conditions (Lynn & Hicks, 2007). Thus, rather than less activation leading to a smaller difference in lure errors in the match condition relative to the mismatch condition, participants may have been better able to use the font a lure was tested in to reject lures when only two associates were studied and the font was used to study its associates. If this were the case, it would also lead to a smaller font-match effect when two associates were studied, but the underlying basis would not be the mechanisms proposed by activation-monitoring theory or global matching models. Experiment 2 tested these alternative interpretations.

Experiment 2

In this experiment, all themes were composed of ten associates related to a single lure. Each theme's study items were presented in two fonts. One of the fonts was used to present eight of the theme's study items, while the other font was used to present the remaining two study items from the same theme. This change in procedure was intended to directly study the influence of testing lures in a font that was associated with many (eight) or few (two) of their associates. Studying ten associates for each theme and manipulating the number of associates from that theme presented in a font does this by ensuring that the only basis for differences in lure errors between the two and eight associates conditions is the number of study items related to a lure that were encoded in a particular font. Further, the fact that each lure was related to ten study items should make it impossible for participants to use recall-to-reject in this experiment, because participants generally seem capable of only using recall-to-reject to reduce errors when there are three or fewer studied items related to a lure (Gallo, 2004).

As was specified prior to Experiment 1, both activation-monitoring theory and global-matching models expect lure errors to be higher when lure items are tested in a font used to study eight of their associates compared to a font used to study two of their associates. Fuzzy-trace theory also predicts this result, but does so because lures that are tested in a font that was presented eight times during encoding will match stronger font gist traces compared to lures tested in a font that was presented twice. Thus, Experiment 2 does not discriminate among the candidate theories, but does serve the empirical purpose of ensuring that the results observed in Experiment 1 are not simply due to scale effects or differences in recall-to-reject across conditions.

Method

Participants

Participants were 48 Middlebury College students who participated as part of an Introduction to Psychology research appreciation requirement or in exchange for \$10 payment.

Materials and Design

Stimuli were 48 themes of 10 items each, constructed in the same manner as those for Experiment 1 (mean associative strength from study item to lure = 0.469; range: 0.026 to 0.960; range of mean associative strength for themes: 0.376 to 0.685). Forty eight of the 64 new items from Experiment 1 were chosen to serve as new items on the recognition memory tests. Ninety-six fonts from Arndt (2006) were chosen to present study and test items. The themes were divided in to two study lists of 24 themes each. Thus, the study list length was 240 words.

There was one manipulation in this experiment: *Number of Associates per Font* (two associates vs. eight associates). Two fonts were used to present each theme during a study list. One font was used to present two of a theme's study items, and a second font was used to present the remaining eight study items from that theme. Assignment of items within a theme to the two and eight associate conditions was determined randomly for each participant. The study items assigned to the two associate condition were presented in succession, and were presented in the first two serial positions, middle two positions (5th and 6th), or final two positions within a theme. Thus, the study items assigned to the eight associates condition were either presented in the first eight serial positions, first four and final four serial positions, or final eight serial positions. A counterbalancing scheme ensured that each theme was assigned to the three orderings equally often across participants.

Test lists were composed of one study item from each theme that was presented in the two associates condition, one study item from each theme that was presented in the eight associates condition, the lure item from each theme, and unrelated new items. Study items included on the memory test were chosen randomly from the two associates and eight associates condition of each theme, and were always tested in the font in which they were studied. Lure items were assigned to one of the Number of Associates per Font conditions. Lures assigned to the two associates condition were tested in the font used to present two of its associates on the study list and lures assigned to the eight associates condition were tested in the font used to present eight of its associates on the study list. A counterbalancing scheme ensured that each theme's lure item was presented equally often in the two associates condition and the eight associates condition across participants. Finally, twenty-four new items were presented to assess baseline false alarm rates when a font was presented eight times on a study list compared to when a font was presented twice. New items were presented in the twenty-four fonts used to test lure items. Thus, the test list was 96 words in length (2 study items $\times$ 24 themes per test list + 24 lures + 24 new items).

Procedure

The procedure was identical to that of Experiment 1.

Results

As with Experiment 1, the results of Experiment 2 were analyzed using both probability “old” judgments as well as d' . Analyses of both measures led to the same conclusion, even though false alarms to new items were reliably higher in the eight associates condition ($M = .12$) than in the two associates condition ($M = .09$; $t(47) = 2.31$, $p = .025$). Therefore, we only present analyses of the probability of “old” judgments. The probability of judging studied items “old”

is presented in the top panel of Figure 3, while the probability of judging lure items “old” is presented in the bottom panel. As is evident in Figure 3, studied item recognition was not influenced by the Number of Associates per Font factor ($t(47) = 0.69, p = .491$), while lure false alarms were higher when tested in a font used to study eight of their associates ($M = .62$) compared to a font used to study only two of their associates ($M = .47; t(47) = 7.17, p < .001$).

Discussion

Experiment 2's results showed the same basic pattern as the font match effects found in Experiment 1 – that testing a lure item in the font used to show many of its associates produced more errors than testing a lure item in the font used to show fewer of its associates. Importantly, this difference in font errors can not be explained by overall differences in lure activation, memory activation, or by differences in recall-to-reject processing between the two associates and eight associates conditions, because lures in both conditions were related to ten study items. Thus, the only difference that could produce error differences between the two conditions was how many of a lure's associates were studied in the font used to present the lure at test.

To this point, both activation-monitoring theory and global-matching models are able to explain the results of both Experiments 1 and 2, while fuzzy-trace theory encounters difficulty explaining some of the data from Experiment 1. The final experiment reported in this paper was designed to discriminate between the explanations of activation-monitoring theory and global-matching models for font-match effects on false recognition, as well as to provide additional tests of fuzzy-trace theory.

Experiment 3

Experiment 3 was a replication and extension of Experiment 2. As in Experiment 2, ten associates were studied for all themes. Further, each theme was presented such that eight of the associates were studied in one font, and the remaining two associates were studied in a second font. At test, some lure items were shown in one of the two fonts used to study their associates, replicating Experiment 2 (the *match* condition). Other lure items were tested in a font used to study a different lure's associates (the *mismatch* condition). Some of the lures in the mismatch condition were tested in a font shown twice during encoding, while other lures were tested in a font shown eight times during encoding.

As was outlined prior to Experiment 2, all three theories predict more errors when lure items are tested in a font used to study eight of their associates compared to a font used to study two of their associates. Further, both activation-monitoring theory and global matching models predict more lure errors in the match condition than in the mismatch condition, and that study-test match will interact with the number of associates studied, while fuzzy-trace theory does not predict those outcomes (see Experiment 1 predictions). In addition, the three theories differ regarding whether the number of times a font was shown during encoding should influence false alarms when a lure is tested in the mismatch condition. Specifically, global-matching models and fuzzy-trace theory predict a difference in false alarms, while activation-monitoring theory does not. The basis for these predictions can be understood by considering each theory's account of study-test match effects for lure items.

Activation-monitoring theory suggests that activation of a lure's representation causes the lure to become associated with encoding context. Thus, the more a lure representation is activated during encoding, the stronger the association between the lure representation and the encoding context (font) of its studied associates. When lure items are tested in the match condition, the lure-font association is cued, allowing the strength of the lure-font association to influence lure errors. However, when a lure item is tested in the mismatch condition, the lure-font association will not serve as a basis for responding, leaving only the activation level of the lure's

representation to support lure errors. Thus, because all lures had ten of their associates presented during encoding, whether a lure item is tested in a font used to study two or eight associates of a different lure should not impact error rates.

Fuzzy-trace theory suggests that lure errors occur because lures match both a semantic gist trace that represents the semantic commonalities of the lure's associates, as well as a font gist trace that represents the commonalities of the font in which the lure is tested. Thus, the theory expects a higher false alarm rate for lure items tested in a font that was shown eight times during encoding because those lures will match a stronger font gist trace compared to lures tested in a font that was shown twice during encoding.

Finally, global-matching models suggest that the influence of presenting lures in the same font as their studied associates occurs because of interactive cueing. Coupling the interactive cueing assumption with global matching allows global-matching models to explain why lure error rates are higher when tested in the font used to present eight of their associates compared with only two of their associates. Specifically, when a lure is tested in the font used to study eight of its associates, that probe will partially match item information for all ten of its associates, and for eight of those memory traces, it will also match the context information in the same memory trace. In comparison, when a lure is tested in the font used to study two of its associates, that probe will partially match item information for all ten of its associates, but only will match context information for two of those ten representations. Thus, testing a lure item in the font used to present eight of its associates will create greater overall memory activation than testing a lure item in the font used to present two of its associates. In contrast, when a lure is tested in the mismatch condition, it activates item and context information in separate memory traces. Each lure probe will partially match ten item traces (those of its associates) as well as context information that was encoded as part of the memory representation of another lure's associates (those that resemble the font in which the lure is tested). Thus, when a lure is tested in a font used to study eight associates of a different lure, it will match eight context traces in memory, and when a lure is tested in a font used to study two associates of a different lure, it will match only two context traces in memory. The result is that memory activation, and therefore lure errors, should be higher when lures are tested in a font used to present eight study items than when they are tested in a font used to present two study items, even when the test font was not used to study the lure's associates.

In sum, both activation-monitoring theory and global-matching models expect lure errors to be higher when they are tested in a font used to show the lures' associates, and that the size of the difference between the match and mismatch conditions will be greater when lures are tested in a font shown eight times, while fuzzy-trace theory does not. Additionally, the theories differ regarding whether lure errors are expected to differ when lures are tested in a font used to study eight or two associates of a different lure (i.e., in the mismatch condition). Activation-monitoring theory predicts no difference in lure errors, while fuzzy-trace theory and global-matching models predict more errors when a lure is tested in a font used to study eight associates of a different lure.

It is possible to test the mismatch condition prediction of fuzzy-trace theory and global-matching models with new item false alarms from Experiments 1 and 2. The basis for the prediction that false alarm rates should be higher when a font was presented eight times compared to two times is the same for new items as it is for lure items in the mismatch condition – a test probe matching stronger font gist traces (fuzzy-trace theory) or a test probe matching a larger number of context traces in memory (global-matching models). Thus, both fuzzy-trace theory and global-matching models predict a difference between false alarm rates for items tested in a font presented eight times compared to two times regardless of whether the test item was related to items that occurred in the study list (i.e., lure items), or was specifically selected

to be unrelated (i.e., new items). In both Experiments 1 and 2, the mean false alarm rate was higher when new items were tested in a font that was presented eight times, although that difference was only reliable in Experiment 2. One plausible reason for the small mean differences, as well as the inconsistent reliability of those differences is that new item false alarms were relatively low in both Experiments 1 and 2. Thus, Experiment 3 provides a stronger test of the predictions of activation-monitoring theory, fuzzy-trace theory, and global-matching models by examining lure false alarm rates, which tend to be much higher than new item false alarm rates.

Computing d' for lures allows one final test of fuzzy-trace theory and global-matching models. Both theories predict that computing d' should eliminate the effects of testing lure items in a font presented eight vs. two times in the mismatch condition. Recall that the reason global-matching models predict that both lures in the mismatch condition and new items will have higher false alarm rates when tested in a font presented eight times during encoding is that those items will match eight context traces in memory, while lures in the mismatch condition and new items tested in a font presented twice during encoding only match two context traces. Similarly, fuzzy-trace theory predicts this difference because lures and new items that are tested in a font that was shown eight times during encoding will match a stronger font gist trace than lures and new items that are tested in a font that was shown twice during encoding. Computing d' for lure items in the mismatch condition uses new item error rates as a measure of the extent to which lure errors are influenced by the memory activation resulting from testing an item in a given font, which removes the influence of a test item matching font gist traces of differing strength (fuzzy-trace theory) or matching different numbers of context traces in memory (global-matching models). Thus, d' for lure items in the mismatch condition assesses how much of lures' memory activation was due to lures matching semantic gist traces (fuzzy-trace theory) or item traces of their studied associates (global-matching models). Because all lure items in this experiment had the same number of associates studied, ten, the strength of semantic gist traces or the amount of memory activation measured by d' should be equivalent in the eight and two font repetitions conditions.

However, the two theories differ regarding how computation of d' should impact lure errors in the match condition. Global-matching models predict that computing d' should not eliminate the difference between lures tested in a font used to study two or eight of their associates (i.e., in the match condition) because of interactive cueing. Interactive cueing ensures that when a lure matches both item and contextual information for eight traces in memory relative to two traces, the activation disparity will be greater than that caused by only matching two or eight context traces in isolation. Thus, the disparity in memory activation between lures tested in the two and eight associates conditions when they are shown in a font used to study their associates (i.e., in the match condition) should be greater than the disparity in memory activation between new items tested in a font shown twice vs. eight times during encoding. As a result, d' for lures in the match condition should still reveal a difference between the two and eight associates conditions.

In contrast, because fuzzy-trace theory assumes that semantic and font gist traces are stored separately, computing d' for lure items should eliminate the difference between lures tested in a font used to study two of their associates and lures tested in a font used to study eight of their associates. The basis for this prediction is the same as that for lure items in the mismatch condition: Computing a measure of discriminability adjusts false alarm rates for the strength of font gist traces that result from studying a font two vs. eight times, leaving only the strength of the semantic gist trace to produce differences between conditions. Because all themes contained ten study items, the strength of the semantic gist trace should be similar in both the eight and two associates studied conditions. Thus, when differences due to font gist representation strength are taken in to account by d' , the difference in false alarm rates between

the eight and two associates studied conditions should be eliminated according to fuzzy-trace theory.

Method

Participants

Participants were 84 Middlebury College students who participated in exchange for \$10 payment or as part of an Introduction to Psychology research appreciation requirement.

Materials and Design

The materials were the same as those used for Experiment 2. Two independent variables were manipulated in Experiment 3: *Number of Associates per Font* (two vs. eight), and *Study-Test Match* (match vs. mismatch). Number of Associates per Font was implemented within study lists in the same way as Experiment 2. Therefore, the study list length was the same, 240 words (24 themes), as it was in Experiment 2. Study-Test Match was implemented by presenting test items either in a font used to present the theme's associates during encoding (the *match* condition) or a font used to present a different theme's associates during encoding (the *mismatch* condition). The match condition replicated Experiment 2, such that studied items in the match condition were tested in the font used to present them during encoding, and lure items were tested in one of the two fonts used to study their associates during encoding – either the font used to study two of its associates or the font used to study eight of its associates. Implementation of the mismatch condition for studied items occurred within each level of the Number of Associates per Font factor. Thus, studied items assigned to the eight associates condition were tested in a font that was presented eight times during the study list, but that was used to show a different theme's study items. Similarly, study items assigned to the two associates condition were tested in a font that was presented twice during the study list, but was used to show a different theme's study items. Similar to Experiment 2, one study item from each of the number of associates per font conditions was randomly selected to be presented on the test list. For lure items in the mismatch condition, those assigned to the eight associates condition were tested in a font used to show eight associates from a theme not related to that lure during study, and lure items assigned to the two associates condition were tested in a font used to show two associates from a theme not related to that lure. Finally, new items were tested in each of the fonts used to test lure items. Thus, new items were tested in fonts assigned to the four experimental conditions created by combining the Number of Associates per Font and Study-Test Match factors. The test list length was therefore 96 words (2 study items $\times$ 24 themes per list + 24 lures + 24 new items). Themes were rotated through the four experimental conditions such that each theme served in each experimental condition equally often across participants.

Procedure

The procedure for Experiment 3 was the same as Experiment 1.

Results

The results from Experiment 3 were analyzed with 2×2 ANOVAs using both probability old judgments and d' as dependent variables. The results from analyzing probability old judgments and d' produced identical conclusions for analyses of studied items, but produced slightly different conclusions for analyses of lure items. Thus, we describe the results of analyses of both dependent measures together when they converge, and then note the critical difference in the lure item results. Analyses of probability old judgments (Figure 4, top panel) and d' (Figure 5, top panel) for studied items indicated that there was a main effect of Study-Test Match ($F(1,83) = 83.04$, $MSE = .012$, $p < .001$ for $p(\text{old})$ analyses; $F(1,83) = 83.36$, $MSE = .310$, $p < .$

001 for d' analyses), indicating that study item memory was better in the match condition than in the mismatch condition. Neither the main effect of Number of Associates per Font ($F(1,83) = 0.67$, $MSE = .003$, $p = .417$ for $p(\text{old})$ analyses; $F(1,83) = 2.78$, $MSE = .224$, $p = .099$ for d' analyses) nor the interaction ($F(1,83) = 1.64$, $MSE = .003$, $p = .204$ for $p(\text{old})$ analyses; $F(1,83) = 0.15$, $MSE = .232$, $p = .696$ for d' analyses) was reliable in the analyses of either dependent measure.

Analyses of probability old judgments (Figure 4, bottom panel) and d' (Figure 5, bottom panel) for lure items both produced reliable main effects for Number of Associates per Font ($F(1,83) = 47.66$, $MSE = .014$, $p < .001$ for $p(\text{old})$ analyses; $F(1,83) = 12.02$, $MSE = .271$, $p = .001$ for d' analyses) and Study-Test Match ($F(1,83) = 37.01$, $MSE = .013$, $p < .001$ for $p(\text{old})$ analyses; $F(1,83) = 14.27$, $MSE = .305$, $p < .001$ for d' analyses), as well as an interaction ($F(1,83) = 8.06$, $MSE = .015$, $p = .006$ for $p(\text{old})$ analyses; $F(1,83) = 4.56$, $MSE = .277$, $p = .036$ for d' analyses). For both dependent variables, the Study-Test Match effect was larger in the eight associates condition than in the two associates condition (Consult Figures 4 & 5). However, the details of the interaction differed for probability old judgments and d' analyses. Specifically, while both lure false alarms (i.e., $p(\text{old})$ judgments; $t(83) = 6.96$, $p < .001$) and d' ($t(83) = 3.90$, $p < .001$) were higher in the eight associates condition when lures' test font *matched* that of their studied associates, the comparison between the two and eight associates conditions when lures' test font *mismatched* that of their associates was reliable for analyses of lure false alarms, but not for analyses of d' . False alarms to lure items were reliably higher in the eight associates condition ($M = .54$; $t(83) = 2.73$, $p = .008$) than in the two associates condition ($M = .49$) when lures' test font mismatched the font in which its associates were studied. However, d' analyses of this same comparison failed to produce a reliable difference between lure errors in the two associates ($M = 1.03$) and eight associates conditions ($M = 1.11$; $t(83) = 0.93$, $p = .353$). The difference in outcomes between these two dependent measures reflects the fact that false alarms to new items were marginally, but not reliably, higher in the eight associates condition ($M = .18$) compared to the two associates condition ($M = .16$; $t(83) = 1.88$, $p = .064$).

Discussion

These results favor global matching models' explanation of DRM lure errors over the explanation offered by fuzzy-trace theory and activation-monitoring theory. Two critical findings from Experiment 3 are predicted by global matching models, but are not readily explained by activation-monitoring theory. First, lure false alarms differed between the two and eight associates conditions when lures were tested in the mismatch condition. Second, there was not a difference in d' between lure errors in the two and eight associates condition when the lure was tested in the mismatch condition, even though all of the other regularities evident in the lure false alarm rate data were maintained in the d' analyses. Thus, the change in reliability between probability old analyses and d' did not occur simply because d' is insensitive to condition differences, but instead was specific to lure d' in the mismatch condition. Critically, this was the only comparison where global-matching models predicted d' computations should change the patterns in the lure error data.

Activation-monitoring theory does not possess a mechanism that can explain lure false alarm rate differences between the two and eight associates conditions when the font in which a lure item was tested did not match the font in which the lure's associates were studied. The basis for lure errors in activation-monitoring theory is how much a lure's representation is activated by the study of its associates, and font match effects are explained as a consequence of associations formed between a font and lure representation during the study of its associates. Thus, there is no provision in activation-monitoring theory for how a font, or other contextual features, can influence lure error rates when the font in which a lure was tested was not

experienced while the lure was active in semantic memory (i.e., as occurs in the mismatch condition).

Although fuzzy-trace theory can account for the data patterns that are problematic for activation-monitoring theory, it nevertheless encounters difficulty explaining two key results from Experiment 3. First, as with the results of Experiment 1, fuzzy-trace theory's proposal that semantic and font gist are stored separately does not offer a mechanism that explains lure error differences between the match and mismatch conditions. Second, fuzzy-trace theory encounters difficulty explaining why computing d' for lures did not eliminate the difference between the eight associates and two associates conditions when the font used to test lure items matched that of its studied associates.

Thus, the results of this experiment illustrate critical differences between the explanation of lure errors offered by activation-monitoring theory, fuzzy-trace theory and global-matching models. Specifically, they point to the importance of both integrated storage of item and context information in memory and interactive cueing for explaining why false memories are enhanced by testing lures in the same encoding context as their studied associates. Similarly, these results point to the importance of test probes that partially match separate representations in memory for explaining why false memories can be enhanced by features that were encoded during a study episode, even when those features were not directly associated with the concept being tested. While activation-monitoring theory and fuzzy-trace theory each contain processes that exhibit the characteristics of one of these mechanisms, only global-matching models possess both mechanisms.

General Discussion

The experiments reported in this paper produced three data patterns of importance. First, lure errors increased when lures were tested in the same font used to study their associates compared to when lures were tested in a font that was studied, but was not used to study their associates (Experiments 1 & 3). Second, the effect of study-test match on lure errors increased as the number of associates studied increased (Experiments 1 & 3). Third, testing a lure in the font used to study the majority of its associates produced more errors compared to testing a lure in the font used to study the minority of its associates (Experiments 2 & 3).

These results reinforce and extend evidence that people experience lure items in the DRM paradigm as authentic, episodic memories (Roediger & McDermott, 1995; Hicks & Hancock, 2002; Hicks & Starns, 2006a). In particular, the lure errors observed in these studies suggest that testing lures in a context that was uniquely associated with its studied associates exacerbated people's belief that lures were experienced, just as reinstating episodic context does for authentic memories (Reder, et al., 2002; Smith, 1979). While context effects in recognition memory tend to be inconsistent (Smith, 1988), studies of perceptual matching from study to test suggest that a critical variable in determining study-test congruence effects is the number of concepts that are associated with a perceptual format (Reder, et al., 2002; Park, Arndt, & Reder, 2006). Thus, when a perceptual format such as a font or voice is associated with relatively few concepts, as was the case in these studies, study-test match effects tend to be larger (and reliable) compared to conditions where a perceptual format is associated with a large number of concepts. In addition, research on context effects in recognition memory suggests that the conditions most likely to enhance recognition memory performance (i.e., discrimination) are those where items and contexts are integrated with one another (Murnane, Phelps, & Malmberg, 1999). Although lures were not necessarily integrated with encoding context in these studies, it is likely that the representations of their studied associates were integrated with context, due to the fact that study items were presented in the fonts that were used as contexts in these studies. Thus, in comparable empirical conditions to those used in

the present studies, authentic memories tend to be improved when encoding context is reinstated, just as the effects of reinstating encoding context exacerbated lure errors. Further, the strength of the memory illusion increased when there were more episodic traces in memory relating a lure's associates to a particular encoding context. These two basic results suggest similar experiential characteristics to those found with studies examining phenomenological judgments of recognition (Roediger & McDermott, 1995) and source attributions for lure items (Hicks & Hancock, 2002; Hicks & Marsh, 2001; Hicks & Starns, 2006a; Roediger, et al, 2004). Thus, these studies, combined with substantial existing evidence, suggest that people experience lures as authentic memories, with many of the same characteristics as true memories.

Of the three theoretical views tested by these studies, only global-matching models were able to consistently explain the results from these experiments; activation-monitoring theory and fuzzy-trace theory both encountered difficulty explaining the entirety of the results. Thus, font-matching effects seem to be most readily explained by the mechanisms inherent in global-matching models. This in turn suggests that participants behave as if lures were episodically experienced because lures are similar to multiple traces in memory, none of which are encoded representations of the lure itself. Thus, these data argue against views suggesting that participants' conviction that lures were studied is an authentic reflection of the state of lure representations in memory (activation-monitoring theory) as well as views suggesting lure items are endorsed with strong conviction based upon memory traces that do not contain detailed contextual information (fuzzy-trace theory). Rather, these data support the view that participants' conviction that lures were studied in a specific format is a consequence of quick, efficient, global-matching processes interacting with multiple memory traces that contain information about the context in which lures' associates were encoded (Hicks & Starns, 2006b).

In addition to providing a comprehensive account of the present data, global-matching models readily accommodate other data that suggest participants believe lure items occurred in a specific encoding context, and do so using their standard assumptions. For example, global matching models can explain why people tend to attribute lures to the source of their studied associates (Roediger, et al., 2004), as well as regularities of lure source attributions such as the source-strength effect (Hicks & Hancock, 2002; Hicks & Starns, 2006a), as detailed by Hicks and Starns (2006b). Although activation-monitoring theory can explain source attributions to lure items, the theory does so by positing an extension of its basic principles where lure representations can be associated with encoding context when they become strongly activated. Thus, parsimony considerations favor global matching models' account of source attributions for lures. Further, the tendency to attribute lures to the source of their strong associates poses difficulty for fuzzy-trace theory's explanation of DRM lure errors. Specifically, the general notion that lure errors are underlain by gist representations does not provide sufficient representational specificity to mediate attributions of lure items to the source of their studied associates (see Hicks & Hancock, 2002 for a similar argument). Rather, the retrieval of gist representations to make lure source attributions predicts random attribution of studied sources to lure items instead of the observed tendency to attribute lures to the source of their strong associates (Hicks & Hancock, 2002). Thus, fuzzy-trace theory's reliance on gist representations as the basis for lure errors in the DRM paradigm may require revision in order to explain data such as those reported in this paper as well as source memory data that demonstrate the specificity of source knowledge participants have for lure items. In sum, not only do global-matching models provide a more comprehensive account of lure source judgments and font-match effects than activation-monitoring theory and fuzzy-trace theory, but they also provide a more parsimonious account.

In addition to explaining source attributions for lure items, global-matching models provide an account of other data suggesting people believe lures were episodically experienced, such as the regularities noted in the introduction. For example, global matching readily explains why lure items show high error rates (Arndt & Hirshman, 1998). Further, global-matching models can explain why participants claim to recollect lures' study presentation. One source for such judgments could be that they are based upon very strong memory signals that render lures' activation of memory comparable to the overall activation of memory caused by probing memory with a studied item (e.g., Donaldson, 1996; Dunn, 2004; Hirshman & Master, 1997). Alternatively, when global-matching models are applied to context effects in recognition, memory probes are assumed to match both item and contextual information in memory. Thus, it is possible that simultaneous activation of item traces of a lure's associates and their associated contextual information supports recollection judgments. Finally, global-matching models seem capable of explaining why people make more lure errors when they are asked to make source attributions than when they are simply asked to make old-new recognition judgments (Hicks & Marsh, 2001). Specifically, searching memory for episodic details, such as source information, is required when participants are completing a source memory task, while searching for source information is not required in order to make an old-new judgment. Thus, both item and context information would be expected to contribute to recognition judgments when participants complete a source memory task, producing greater overall activation of memory and more lure errors compared to old-new recognition, where item information's contribution alone is likely to be the dominant basis for memory activation. Consequently, not only can global matching models accommodate the present results, but they also provide an account of several major regularities that suggest participants believe lures were experienced in a particular encoding context. Therefore, the present results, as well as broad trends in the existing literature, suggest that global-matching models offer a viable account of false memory errors in the DRM paradigm (see also Kimball, Smith, & Kahana, 2007 for application of a global memory model, SAM, to false recall).

In closing, this research highlights two issues that are important to consider in future work examining the mechanisms underlying false memories. First, although much of the research in the DRM paradigm has been designed to test activation-monitoring theory and fuzzy-trace theory, future research may find it profitable to also consider global-matching models as an explanatory framework worthy of testing. For example, it may prove useful to explore how well single-process global-matching models can account for data that have been argued to provide support for the dual-process accounts of memory errors proposed by activation-monitoring theory and fuzzy-trace theory (e.g., Arndt & Gould, 2006) in much the same way that testing single- and dual-process models has aided understanding of recognition memory in general (Diana, Reder, Arndt, & Park, 2006). Similarly, explicitly juxtaposing the accounts of false memory phenomena offered by theories that make disparate assumptions, as was done in the present paper, holds the potential to stimulate research that furthers understanding of the mechanisms by which false memories arise, which in turn may inform techniques for minimizing memory errors. Thus, explicitly comparing the account of memory errors provided by global-matching models with those of activation-monitoring theory and fuzzy-trace theory may further understanding of the extent to which different types of representations and retrieval processes operate to produce and limit false memories. Second, employing specific visual contexts, such as the fonts used in these studies, offers a straightforward method for examining the sensitivity of lure errors to manipulations of encoding context. As evidenced in the first and third experiments reported here, the use of fonts in the present studies afforded the opportunity to correlate each theme with a specific visual context, which in turn enabled the specificity of lure errors to be probed with great precision. This precision ultimately proved critical to distinguishing between the theories tested by these experiments, as well as illustrating the specificity of contextual knowledge memory retrieval provides when lure items are tested. Thus, manipulating contextual overlap between the encoding of a lure's associates and the lure's

test presentation can complement studies that use phenomenological judgments (Roediger & McDermott, 1995) and source memory judgments (Hicks & Hancock, 2002; Hicks & Starns, 2006b; Roediger, et al., 2004) to improve understanding of the representational characteristics of false memories.

Acknowledgments

I thank Kelly Bennion, Sophie Dorot, Emer Feighery, Erin Frazier, Jenny Galgano, Nitzah Gebhard, Molly Huff, Jeff Lam, Ellie Molyneux, Yina Ng, Ashley Pfaff, & Emily Read for their work completing these studies.

This research was supported by grant 1R15 MH077665 from the National Institutes of Health.

References

- Arndt J. Distinctive information and false recognition: The contribution of encoding and retrieval factors. *Journal of Memory and Language* 2006;54:113–130.
- Arndt J, Gould C. An examination of two-process theories of false recognition. *Memory* 2006;14:814–833. [PubMed: 16938694]
- Arndt J, Hirshman E. True and false recognition in MINERVA2: Explanations from a global-matching perspective. *Journal of Memory and Language* 1998;39:371–391.
- Arndt J, Reder LM. The effect of distinctive visual information on false recognition. *Journal of Memory and Language* 2003;48:1–15. [PubMed: 19079562]
- Brainerd, CJ.; Reyna, VF. *The Science of False Memory*. New York: Oxford University Press; 2005.
- Brainerd CJ, Reyna VF, Kneer R. False recognition reversal: When similarity is distinctive. *Journal of Memory and Language* 1995;34:157–185.
- Brainerd CJ, Wright R, Reyna VF, Mojardin AH. Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2001;27:307–327.
- Clark SE, Gronlund SD. Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review* 1996;3:37–60.
- Deese J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology* 1959;58:17–22. [PubMed: 13664879]
- Diana RA, Reder LM, Arndt J, Park H. Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review* 2006;13:1–21. [PubMed: 16724763]
- Donaldson W. The role of decision processes in remembering and knowing. *Memory & Cognition* 1996;24:523–533.
- Dunn JC. Remember-Know: A Matter of Confidence. *Psychological Review* 2004;111:524–542. [PubMed: 15065921]
- Gallo DA. Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30:120–128.
- Geiselman RE, Fisher RP, Cohen G, Holland H, Surtes L. Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science & Administration* 1986;14:31–39.
- Geiselman RE, Fisher RP, Firstenberg I, Hutton LA, Sullivan SJ, Avetissian IV, Prosk AL. Enhancement of eyewitness memory: An empirical evaluation of the Cognitive Interview. *Journal of Police Science and Administration* 1984;12:130–138.
- Hicks JL, Hancock TW. Backward associative strength determines source attributions given to false memories. *Psychonomic Bulletin & Review* 2002;9:807–815. [PubMed: 12613687]
- Hicks JL, Marsh RL. False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2001;27:375–383.
- Hicks JL, Starns JJ. The roles of associative strength and source memorability in the contextualization of false memory. *Journal of Memory and Language* 2006a;54:39–54.

- Hicks JL, Starns JJ. Remembering Source Evidence From Associatively Related Items: Explanations From a Global Matching Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2006b;32:1164–1173.
- Hintzman D. Judgments of frequency and recognition memory in a multiple trace memory model. *Psychological Review* 1988;95:528–551.
- Hirshman E, Master S. Modeling the conscious correlates of recognition memory: Reflection on the remember-know paradigm. *Memory and Cognition* 1997;25:345–351.
- Jones TC, Jacoby LL. Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language* 2001;45:82–102.
- Kimball DR, Smith TA, Kahana MJ. The fSAM model of false recall. *Psychological Review* 2007;114:954–993. [PubMed: 17907869]
- Kucera, H.; Francis, WN. Computational analysis of present-day American English. Providence, RI: Brown Univ. Press; 1967.
- Lindsay DS, Johnson MK. The eyewitness suggestibility effect and memory for source. *Memory & Cognition* 1989;17:349–358.
- Loftus EF. The reality of repressed memories. *American Psychologist* 1993;48:518–537. [PubMed: 8507050]
- Loftus GR. On interpretation of interactions. *Memory & Cognition* 1978;6:312–319.
- Lynn, SD.; Hicks, JL. Font reinstatement encourages, and sometimes discourages, false recognition. Poster presented at the Annual Meeting of the Psychonomic Society; Long Beach, CA. 2007.
- Murnane K, Phelps MP, Malmberg K. Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General* 1999;128:403–415. [PubMed: 10650581]
- Nelson, DL.; McEvoy, CL.; Schreiber, TA. The University of South Florida word association, rhyme, and word fragment norms. 1998. Available from <http://www.usf.edu/FreeAssociation/>
- Park H, Arndt J, Reder LM. A contextual interference account of distinctiveness effects in recognition. *Memory & Cognition* 2006;34:743–751.
- Reder LM, Donavos D, Erickson MA. Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition* 2002;30:312–323.
- Reyna VF, Lloyd F. Theories of false memory in children and adults. *Learning and Individual Differences* 1997;9:95–123.
- Robinson KJ, Roediger HL III. Associative processes in false recall and false recognition. *Psychological Science* 1997;8:231–237.
- Roediger HL III, McDermott KB. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1995;21:803–814.
- Roediger HL III, McDermott KB, Pisoni DP, Gallo DA. Illusory recollection of voices. *Memory* 2004;12:586–602. [PubMed: 15615317]
- Roediger HL III, Watson JM, McDermott KB, Gallo DA. Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review* 2001;8:385–407. [PubMed: 11700893]
- Schacter, DL. Memory distortions: How minds, brains, and societies reconstruct the past. Cambridge, MA: Harvard University Press; 1995.
- Shiffrin RM, Steyvers M. A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review* 1997;4:145–166.
- Slotnick SD, Schacter DL. A sensory signal that distinguishes true from false memories. *Nature Neuroscience* 2004;7:664–672.
- Smith SM. Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory* 1979;5:460–471.
- Smith, SM. Environmental context-dependent memory. In: Davies, GM.; Thomson, DM., editors. *Memory in context: Context in memory*. New York: Wiley; 1988. p. 13-34.
- Tulving E. Memory and consciousness. *Canadian Psychologist* 1985;26:1–12.
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language* 2002;46:441–517.
- Zaragoza MS, Lane SM. Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994;20:934–945.

CHIRP
BUMPER
FLOOD
GENIUS
HANGER
STING

Figure 1.
Examples of fonts used to present stimuli.

Figure 2. Proportion of old responses in Experiment 1 for studied items (top panel) and lure items (bottom panel) as a function of Number of Associates Studied and Study-Test Match. Error bars indicate the standard error of the mean.

Figure 3. Proportion of old responses in Experiment 2 for studied items (top panel) and lure items (bottom panel) as a function of Number of Associates per Font. Error bars indicate the standard error of the mean.

Figure 4. Proportion of old responses in Experiment 3 for studied items (top panel) and lure items (bottom panel) as a function of Number of Associates per Font and Study-Test Match. Error bars indicate the standard error of the mean.

Figure 5. d' for studied items (top panel) and lure items (bottom panel) in Experiment 3 as a function of Number of Associates per Font and Study-Test Match. Error bars indicate the standard error of the mean.