Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 2, 391–404

0278-7393/11/$12.00 DOI: 10.1037/a0025548

Illusory Expectations Can Affect Retrieval-Monitoring Accuracy

Ian M. McDonough
University of Texas at Dallas

David A. Gallo
University of Chicago

The present study investigated how expectations, even when illusory, can affect the accuracy of memory decisions. Participants studied words presented in large or small font for subsequent memory tests. Replicating prior work, judgments of learning indicated that participants expected to remember large words better than small words, even though memory for these words was equivalent on a standard test of recognition memory and subjective judgments. Critically, we also included tests that instructed participants to selectively search memory for either large or small words, thereby allowing different memorial expectations to contribute to performance. On these tests we found reduced false recognition when searching memory for large words relative to small words, such that the size illusion paradoxically affected accuracy measures (d' scores) in the absence of actual memory differences. Additional evidence for the role of illusory expectations was that (a) the accuracy effect was obtained only when participants searched memory for the aspect of the stimuli corresponding to illusory expectations (size instead of color) and (b) the accuracy effect was eliminated on a forced-choice test that prevented the influence of memorial expectations. These findings demonstrate the critical role of memorial expectations in the retrieval-monitoring process.

Keywords: false memory, expectations, retrieval monitoring, metacognition, distinctiveness heuristic

It has long been assumed that episodic memory is affected by expectations. Because memory can be an imperfect and reconstructive process, people use expectations to help decide what likely occurred in the past (e.g., Johnson, Hashtroudi, & Lindsay, 1993; Schacter, Norman, & Koutstaal, 1998). While early research focused on how expectations can distort memory (e.g., Bartlett, 1932; for review, see Roediger, 1996), more recent research has focused on how expectations can improve memory accuracy through retrieval-monitoring processes (Schacter & Wiseman, 2006). For example, Schacter, Israel, and Racine (1999) proposed a retrieval monitoring process dubbed *the distinctiveness heuristic* to explain why testing memory for more-distinctive recollections (e.g., pictures) leads to less false recognition on a recognition memory test compared with less-distinctive recollections (e.g., words). According to this theory, more perceptual details are expected to be recollected for pictures than for words, leading to more conservative response criteria when searching memory for pictures (e.g., "This item probably wasn't presented as a picture, because I'd remember that"). In contrast, when participants are tested for words, more liberal response criteria would be used because most of the items likely elicit vague recollections or feelings of familiarity rather than distinctive recollections.

The effect of many other encoding manipulations on episodic memory errors also has been attributed to memorial expectations, suggesting that such expectations may play a central role in the decision processes used at retrieval. These manipulations have included visual compared with auditory presentation (e.g., Pierce & Gallo, 2011; Smith & Hunt, 1998), deep compared with shallow processing (e.g., Gallo, Meadow, Johnson, & Foster, 2008), and autobiographical compared with semantic elaboration (e.g., McDonough & Gallo, 2008). Memorial expectations also have been argued to contribute to the false recognition portion of the mirror effect—when studying one class of items leads to more hits and less false recognition relative to another class of items—such as expecting to recollect low-frequency words more than high-frequency words (e.g., Cary & Reder, 2003; Stretch & Wixted, 1998; for review, see Greene, 2007). Finally, memorial expectations are thought to play a central role in source memory tests (Johnson et al., 1993; Johnson & Raye, 1981). According to the source-monitoring framework, people decide whether an item occurred in a potential source depending on their memorial expectations for the features that are characteristic of that source. Although different theories have been proposed to explain these various findings, one common theme is that memory errors can be avoided by relying on one's retrieval expectations and resulting response criteria, a general process that has been described as diagnostic monitoring (see Gallo, 2010).

Although distinctiveness effects on false recognition have been attributed to differences in memorial expectations, another potentially important factor that has been relatively overlooked in these

This article was published Online First September 26, 2011.

Ian M. McDonough, Center for Vital Longevity, School of Behavioral and Brain Sciences, University of Texas at Dallas; David A. Gallo, Department of Psychology and Center for Cognitive and Social Neuroscience, University of Chicago.

Aspects of these data were presented at the 81st Annual Meeting of the Midwestern Psychological Association (2009, Chicago, Illinois) and at the 50th Annual Meeting of the Psychonomic Society (2009, Boston, Massachusetts). We thank Candace (Hope) Bias, Lucas Tian, Melissa Mongrella, Meghan Putty, and Stephanie Gutierrez for assistance collecting data. We also thank members of the Memory Research Lab for helpful comments on previous versions of this article.

Correspondence concerning this article should be addressed to Ian M. McDonough, Center for Vital Longevity, University of Texas at Dallas, 1600 Viceroy Drive, Suite 800, Dallas, TX 75235. E-mail: ian.mcdonough@gmail.com

studies is differences in feature overlap. For example, on a source memory test, more overlapping features between the sources can lead to more source confusions for studied items (e.g., Goff & Roediger, 1998; Henkel, Johnson, & De Leonardis, 1998; Lyle, Bloise, & Johnson, 2006). Similarly, when one attempts to discriminate between items that did and did not occur on a recognition memory test (old and new items), more overlapping features between the items can lead to more false recognition, consistent with feature-based models of memory (e.g., Chappell & McClelland, 1998; Criss, 2006). One way that overlapping features may lead to false recognition is by creating a sense of familiarity that the nonstudied item has been studied in the absence of any recollected details. Overlapping features also might trigger the retrieval of partial features from studied items, potentially leading to a false recollection (Lampinen, Meier, Arnal, & Leding, 2005). In a sense, these effects of feature overlap can be considered an example of the encoding specificity principle (Tulving & Thompson, 1973), whereby nonstudied items act as retrieval cues for similar features in studied items, thereby enhancing false recognition.

Differences in feature overlap might provide an alternative explanation for the false recognition effects that have been attributed to differences in distinctiveness and corresponding expectations. According to this idea, nonstudied items may be less likely to produce familiarity when tested among more-distinctive studied items compared with less-distinctive studied items, because the nonstudied items are less likely to share common features with distinctive studied items (cf. Chappell & McClelland, 1998; Criss, 2006). For example, distinctive items such as pictures are thought to have fewer overlapping features in memory compared with less-distinctive items such as words (e.g., Moscovitch & Craik, 1976; Nelson, Reed, & McEvoy, 1977; see also Cleary, Morris, & Langley, 2007; Greene, 2004). As a result, nonstudied items presented on a recognition test may be less familiar or less likely to trigger a false recollection when tested for picture memories relative to word memories, leading to decreases in false recognition. Although distinctive items might have richer or more-elaborate features than less-distinctive items, these features may be more tightly bound to each unique representation (e.g., Winograd, Karchmer, & Russell, 1971), making them less likely to generate misleading information at test. Critically, these feature-overlap effects might be triggered automatically, thereby affecting false recognition independent of more-explicit changes in retrieval monitoring or response criteria that may occur (such as the distinctiveness heuristic).

These considerations highlight a potential limitation of prior research on the distinctiveness heuristic, and more generally, they highlight the potential limitation of using false recognition effects or source memory biases to conclude that memorial expectations affect memory accuracy. All prior studies investigating the distinctiveness heuristic have manipulated encoding factors to affect recollection quality and corresponding memorial expectations, but these same manipulations also might have affected feature overlap (for similar ideas, see Verde, Stone, Hatch, & Schnall, 2010). Thus, although differences in retrieval expectations may well have caused the observed false recognition differences, via a controlled retrieval monitoring or criteria-setting process, it also may be that retrieval expectations played no causal role in the creation of false recognition errors. These errors instead may have been driven entirely by relatively automatic feature-overlap effects that were driven by characteristics of the stimuli themselves, with differences in retrieval expectations being formed after the fact, as the result (not the cause) of these differences in memory confusions. More generally, researchers in this area may have overestimated the effects of retrieval monitoring while unintentionally slighting the effects of stimulus control on memory performance (cf. Watkins, 1991).

At the methodological level, the potential effect of feature overlap on false recognition errors typically would be greater when distinctiveness is manipulated between-subjects compared with within–subjects, as in much of the earlier work on the distinctiveness heuristic (Dodson & Schacter, 2002; Schacter et al., 1999). In between-subjects designs, word conditions would be associated with more overall feature overlap than would picture conditions, whereas in within-subject designs, both words and pictures would be studied in the same list, yielding a constant degree of feature overlap.

Less obvious, though, is the potential for the feature-overlap effect to vary across test conditions in within-subject designs, at least when retrieval orientation is experimentally manipulated. For example, Gallo, Weiss, and Schacter (2004) developed the criterial recollection task to more explicitly manipulate retrieval orientation at test, relative to prior studies of the distinctiveness heuristic (which used standard *Yes/No* recognition tests). In their study, the type of studied items (pictures and words) was manipulated within-subject at encoding, but retrieval was selectively oriented toward one class of studied items (pictures) or the other (words) across different test blocks via instructions. Even though the same kinds of items were presented on each test block, false recognition was reduced on the picture test compared with the word test (for related results using misleading instructions, see Dodson & Schacter, 2002). Although these effects were attributed to differences in memorial expectations and response criteria, the point we wish to raise here is that the test conditions also may have differed in terms of feature overlap. When retrieval was selectively oriented toward a specific class of items, the effects of feature overlap might have been specific to the features of that particular search set (cf. Nairne, 2006; Rugg & Wilding, 2000). As a result, the nontarget items may have been more likely to elicit a sense of familiarity or false recollection on the word test than on the picture test due to greater feature overlap with the search set, independent of any changes in response criteria that may have occurred. This possibility highlights the difficulty in separating retrieval expectation effects (and corresponding changes in response criteria) from more-automatic effects of feature overlap on false recognition.

Despite this theoretical ambiguity in prior studies of the distinctiveness heuristic, a study by Westerman, Miller, and Lloyd (2003, Experiment 2) provided some evidence that memorial expectations might affect false recognition independent of feature-overlap effects. Participants were told that they would be studying a subliminal set of either words or pictures when, in fact, only visual noise was presented at study. On a subsequent memory test, participants were presented with words and instructed to endorse those words that felt familiar because they may have been presented at study. Critically, some of the words were primed with a matching word to elicit familiarity (cf. Jacoby & Whitehouse, 1989; Whittlesea, 1993), and these words were falsely recognized more often if participants believed they previously studied words compared with pictures. This false recognition difference was attributed to memorial expectations, whereby participants expected that prime-induced fluency would be more diagnostic of prior presentation in the word condition than in the picture condition. Because no items were actually studied, these results cannot be attributed to feature overlap but instead implicate a role for retrieval expectations. It is unclear whether this expectation effect is analogous to the distinctiveness heuristic described in more typical recognition memory tests or if this effect was unique to the demands established by the test prime procedure (e.g., Gallo, Perlmutter, Moore, & Schacter, 2008). Nevertheless, this study demonstrates how the experimental manipulation of illusory expectations can potentially help to separate retrieval monitoring effects from feature-overlap effects on false recognition.

Current Experiments

The aim of the current experiments was to more definitively determine the extent that expectations can affect false recognition independent of differences in feature overlap (or memory quality), using procedures that are analogous to those of previous studies on the distinctiveness heuristic. To accomplish this aim we used a memorability illusion initially reported by Rhodes and Castel (2008). Rhodes and Castel showed that words presented in large font were predicted to be more memorable than words presented in small font, but subsequent recall did not differ as a function of font size. In another study, they extended this illusion to hearing words in a loud compared with a soft volume (Rhodes & Castel, 2009). Based on these illusions, they argued that people erroneously believe that items processed more easily or fluently should be better remembered.

In the current study we reasoned that manipulating font size should not lead to differences in memory quality or distinctiveness, so that large and small words would elicit the same degree of feature overlap. However, the size manipulation should affect memorial expectations. If expectations alone are sufficient to reduce false recognition, then false recognition should be affected by instructions to selectively search memory for items of one font size compared with the other. Because participants should expect to remember large words better than small words, they should also adopt more conservative response criteria when searching memory for large words, thereby suppressing false recognition via a monitoring process akin to a distinctiveness heuristic. Experiment 1 tested this memorial-expectation hypothesis, and Experiments 2 and 3 provided additional evidence that our results were based on memorial expectations, independent of any potential differences in memory quality (and corresponding feature overlap) between large and small words.

Experiment 1

To explore the role that expectations play in retrieval monitoring, we used the criterial recollection task (Gallo et al., 2004). At study, participants saw words in different font sizes (large and small) and rated each for how confidently they would later remember those words (i.e., judgments of learning [JOLs]). To anticipate, these JOLs confirmed participants' mistaken expectations that large words would be better remembered than small words, replicating Rhodes and Castel (2008). Following the study phase, participants were given memory tests using neutral-sized words as retrieval cues. The same types of items occurred on each of these tests, but we selectively probed memory for one font size or the other by varying the test instructions across testing blocks (i.e., "Did you study this word in large font?" on a large-font test [hereafter called a large test] and "Did you study this word in small font?" on a small-font test [hereafter called a small test]). Importantly, some words had been studied in both large and small font, so that the two sources were not mutually exclusive. Thus, to respond accurately, participants had to selectively search memory for the to-be-recollected (or criterial) information, altering their memorial expectations accordingly.

If participants expected to remember large words better than small words, then they should have used more conservative response criteria on the large test compared with the small test, thereby reducing false recognition of words that were not studied in the criterial format. By keeping the test items the same and varying only the retrieval orientation (via instructions), this method provided a strong test of the memorial-expectation hypothesis. Note that this expectation-based change in response criteria could theoretically affect all of the item types at test, but prior work has suggested that the effect might be most pronounced for highly familiar lures, such as those that were studied but in the incorrect source (cf. Dodson & Schacter, 2002; Gallo et al., 2004; McDonough & Gallo, 2008). Our working assumption was that studied lures elicit a strong sense of familiarity or noncriterial recollection (by design) and hence might be more likely to require additional retrieval-monitoring processes to reject (i.e., searching memory for additional evidence and comparing this evidence with one's criteria, including one's memorial expectations). In contrast, targets can be accepted based on recollection instead of familiarity (cf. Yonelinas, 2002), potentially making these types of items less sensitive to the retrieval-monitoring process under investigation here. Similarly, nonstudied lures can be rejected based on the relative absence of either recollection or familiarity, although these items do sometimes exhibit false recognition suppression effects, potentially depending on their relative level of familiarity.

We also included several additional tests to confirm our assumption that large words were not better remembered than small words (i.e., that the two classes were matched in terms of memory quality or feature overlap). Following the criterial recollection tests, participants received a standard *Old/New* recognition test to measure overall memory strength, along with subjective judgments to estimate recollection, familiarity, and distinctiveness (Gallo, Meadow, et al., 2008; McDonough & Gallo, 2008, 2010). We predicted no significant differences in any of these more standard memory measures, and to anticipate, none were found.

Method

Participants. Thirty-seven students (M_{age} = 20.80 years, SD = 3.12) at the University of Chicago participated in the experiment for course credit or $10. Data from one participant were removed because the participant failed to complete the experiment, resulting in 36 participants.

Materials and design. The current experiment consisted of a 2 (font size: large, small) × 3 (test type: large, small, recollection quality) × 4 (word type: large, small, "both," new) within-subject design. The stimuli included 192 common object words (e.g., *fish,* *hat, bus*). At study, participants viewed words in either large (125-point) or small (25-point) Times New Roman font. Of the 192 words, participants viewed 48 in large font only, 48 in small font only, and 48 in both large and small font ("both" words), and they never studied an additional set of 48 words. Words were counterbalanced across the four types (large-only, small-only, "both," and new). As with previous experiments using the criterial recollection task, "both" words were included at study so that the sources were not mutually exclusive, thus preventing the use of a recall-to-reject strategy (see Gallo, Cotel, Moore, & Schacter, 2007). To ensure that the number of presentations could not be used to detect these "both" words, we repeated the large-only and small-only words, so that every studied word was presented twice.

After the study phase, participants were given three memory tests (large test, small test, and standard recognition memory test). The order of the large and small tests was counterbalanced to allow direct comparisons. The standard recognition memory test always occurred last, so that it would not influence the other two tests. Each test consisted of 64 words (randomly arranged) made of 16 of each type (large-only, small-only, "both," and new). Participants were presented with each word in 75-point Georgia font. Each word at test was presented with the appropriate test prompt (e.g., "Did you study this word in large font?" during the large test). Targets were words that corresponded to the appropriate test (e.g., large-only and "both" words on the large test), studied lures were words that did not correspond to the appropriate test (e.g., small words on the large test), and nonstudied lures were words that were not previously studied. The standard recognition memory test consisted of up to three judgments. The first was an *Old/New* recognition judgment, the second was an *Actually Recollect* or *Very Familiar* judgment analogous to *remember/know* judgments (e.g., Gardiner & Java, 1991; Rajaram, 1993; Tulving, 1985; Yonelinas, 2002), and the third was a unique recollected detail judgment intended to gather subjective evidence of memorial distinctiveness (Gallo, Meadow, et al., 2008; McDonough & Gallo, 2008, 2010). On this final test, participants were given the appropriate prompt for each judgment (i.e., "Did you study this word?" "Actually recollect or very familiar," and "Amount of unique recollected details").

Procedure. During the study phase, participants viewed words one a time in either large or small font (randomly arranged) for 2 s. After 2 s, the word disappeared and a prompt appeared asking how likely participants thought they would remember the word on a scale from 0 (*Definitely Will Not*) to 9 (*Definitely Will*). These JOLs were self-paced and were followed by the next word. Participants were told that their memories would be tested on an unspecified memory test.

Immediately after the study phase, participants took either the large or small test. On the large test, participants were instructed to press *yes* if they remembered viewing the word in large font (large-only or "both" words), otherwise *no* (small-only or new words). Participants were also told that remembering a word that was viewed in small font was irrelevant on this test because, for some words, they viewed the word in both large and small font. They were then asked to make confidence judgments on a scale from 1 (*Low*) to 7 (*High*). Analogous instructions were given for the small test, with the only change being in the to-be-recollected (or criterial) information.

The last test was always the standard recognition memory test. On this test, participants were instructed to press *Yes* if they had previously viewed the word (large-only, small-only, or "both" words), otherwise *No* (new words). If participants responded *Yes*, they were presented with the *Actually Recollect/Very Familiar* judgment. On the *Actually Recollect/Very Familiar* judgment, participants were instructed to press the *1* button if they could recall a specific memory of the test word's presentation (i.e., *Actually Recollect*) and the *2* button if the test word was familiar and they thought it had been presented but could not actually recollect any details about the word (i.e., *Very Familiar*). Participants were also told that the difference between *Actually Recollect* and *Very Familiar* was not confidence. If participants responded *Actually Recollect*, they were presented with the unique recollected details judgment. For this judgment, participants were told to rate the amount of unique details from 1 (*Few Unique Details*) to 7 (*Many Unique Details*). Unique recollected details were defined as details that were different from other words in the study phase, independent of how strong or vivid those recollections may be.

Results and Discussion

All results reported were considered significant at a $p < .05$. Corresponding effect sizes were calculated using Cohen's d for t tests and η_p^2 for analyses of variance (ANOVAs).

Study ratings. Results from the JOLs can be found in Table 1. Consistent with prior work, large words had higher JOL ratings than did small words (Rhodes & Castel, 2008). In addition, words seen a second time had higher JOL ratings than when they were seen the first time. JOLs were entered in a 2 (presentation: first, second) × 2 (size: large, small) repeated-measures ANOVA confirming a main effect of presentation, $F(1, 35) = 8.23$, $MSE = 1.19$, $p = .007$, $\eta_p^2 = .19$; a main effect of size, $F(1, 35) = 34.35$, $MSE = 0.16$, $p < .001$, $\eta_p^2 = .50$; and a Presentation × Size interaction, $F(1, 35) = 8.32$, $MSE = 0.051$, $p = .007$, $\eta_p^2 = .19$. The interaction indicated that the difference between large and small JOL ratings was greater when seen the first time than the second.

Criterial recollection test. Results from the large and small tests can be found in Table 2. We first describe the results within each test, which replicated prior work with this task (all $ps < .001$). Hits to criterial targets (e.g., words that were studied in large font on the large test) were greater than false recognition to studied lures (e.g., words studied in small font on the large test). These results demonstrated that participants had adopted the appropriate retrieval orientation on each test. Second, false recognition to studied lures was greater than false recognition to nonstudied lures, demonstrating that recollection of size information was not perfect and that participants made source memory confusions. Finally, the recognition of words that were studied in one format (large hits and small hits) was greater than the recognition of words that were studied in both formats ("both" hits). This finding stems from the fact that the former were studied twice in the criterial size, whereas the latter were only studied once in the criterial size.

The critical comparison was the level of false recognition to the studied lures across the two tests. As predicted by the memorial-expectation hypothesis, false recognition to studied lures on the large test was lower than those on the small test, $t(35) = 2.31$, $SEM = .056$, $p = .027$, $d = 0.52$. This effect is consistent with the idea that memorial expectations can affect false recognition, be-

Table 1
Mean Judgments of Learning as a Function of Font Size in Experiments 1–3

	Seen one time				Seen two times			
	Large		Small		Large		Small	
Experiment	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
1	5.26	0.23	4.77	0.23	5.66	0.25	5.40	0.26
2								
Size-oriented	5.17	0.24	4.51	0.23	5.34	0.28	4.88	0.30
Color-oriented	5.53	0.18	4.41	0.16	5.52	0.19	4.63	0.19
3								
CR test	5.20	0.22	4.52	0.23	5.82	0.29	5.35	0.30
2AFC test	5.20	0.19	4.39	0.19	5.40	0.29	4.96	0.28

Note. Judgments of learning were made on a scale of 0 (*Definitely Will Not Remember*) to 9 (*Definitely Will Remember*). Large = large-font words; Small = small-font words; *SEM* = standard error of the mean; CR = criterial recollection; 2AFC = two-alternative forced choice.

cause participants were more likely to expect stronger memories for large compared with small words (as reflected in the JOLs). There were no differences across tests between hits to criterial targets ($p = .63$) or false recognition to nonstudied lures ($p = .62$), although there was a trend for fewer hits to "both" targets on the large test relative to the small test ($p = .065$).[1]

We also calculated signal detection measures of discrimination (d') to summarize performance on our criterial recollection tests, corrected for undefined values[2] (see Table 3). Two different d' measures were calculated because there were two kinds of lures on each criterial recollection test, those that were studied in the noncriterial format and those that were nonstudied. The first (d'-studied) compared criterial hits and false recognition with studied lures, whereas the second (d'-nonstudied) compared criterial hits and false recognition with nonstudied lures (for simplicity, "both" items studied in each size format were excluded in this analysis). Although these measures are typically used to determine the strength of memories for studied items, it is important to recognize that d' can be affected by differences in responding to either targets or lures. Under the current experimental conditions, we expected differences in d' owing to the differences that we observed in false recognition to lures, even though we did not expect any actual differences in target memory quality. Consistent with our predictions, d'-studied was greater on the large test than small test when measured using false recognition to studied lures, $t(35) = 2.93$, $SEM = 0.16$, $p = .006$, $d = 0.50$. This effect was not significant when d' was measured using false recognition to nonstudied lures ($p = .25$).

We also analyzed confidence judgments for each correct response between the large and small test. Numerically, confidence judgments were greater on the large test for all words (hits to "both" targets $M = 5.42$, $SE = 0.20$; hits to criterial targets $M = 5.82$, $SE = 0.19$; correct rejections to studied lures $M = 4.19$, $SE = 0.21$; correct rejections to nonstudied lures $M = 5.57$, $SE = 0.20$) than those on the small test (hits to "both" targets $M = 5.04$, $SE = 0.18$; hits to criterial targets $M = 5.37$, $SE = 0.18$; correct rejections to studied lures $M = 4.00$, $SE = 0.21$; correct rejections to nonstudied lures $M = 5.43$, $SD = 0.21$), but confidence ratings only were significantly greater for hits to "both" targets, $t(34) = 2.26$, $SEM = 0.18$, $p = .031$, $d = 0.36$, and hits to criterial targets on the large test, $t(34) = 3.04$, $SEM = 0.15$, $p = .004$, $d = 0.43$, compared with those on the small test. Differences in degrees of freedom are due to one participant not having any correct acceptances on the small test. There were no significant differences in confidence ratings for correct rejections to lures (all $ps > .22$). These differences are consistent with the idea that participants carried a false expectation that large words were better remembered than small words.

Standard recognition memory test. As expected, there were no differences found between large and small words for any of the measures on the standard recognition memory test (see Table 4). Large words did not differ from small words for *Old/New* recognition hits ($p = .40$), *Actually Recollect* judgments ($p = .36$), *Very Familiar* judgments ($p = .44$), familiarity estimates using the *independent remember/know* procedure[3] ($p = .11$), or *Amount of Unique Recollected Details* ratings ($p = .68$). We also calculated d' using studied hits for either large or small words and false recognition to nonstudied words, and as expected, this measure also failed to show a significant difference ($p = .95$).

To summarize the results of Experiment 1, the JOLs obtained at encoding indicated that participants expected to remember large words better than small words, but no differences in actual memory for large words compared with small words were found on any of the standard memory measures. These results confirm the mem-

[1] As described in the introduction, different response criteria might affect targets to the extent that these items (like studied lures) can sometimes elicit relatively vague memories (i.e., a strong sense of familiarity without recollection) that may require additional retrieval-monitoring processes. However, we did not find significant target effects in any of our other comparisons in this experiment or in subsequent experiments.

[2] To avoid undefined values during these types of calculations, we applied a correction proposed by Snodgrass and Corwin (1988), which adds 0.5 to each count and 1.0 to the total number of items.

[3] The *independent remember/know* procedure (see Yonelinas, 2002) transforms the raw proportion of *Very Familiar* judgments as a proportion of only those trials where such a judgment could have been made (1 – *Actually Recollect*). This procedure assumes that recollection and familiarity are independent processes that both contribute to recognition judgments. Because these calculations can result in undefined values, we applied a correction proposed by Snodgrass and Corwin (1988) only when calculating the familiarity estimates, adding 0.5 to each count and 1.0 to the total number of items.

Table 2
Mean Proportion of Items Recognized on Each Criterial Recollection Test in Experiments 1–3

	Experiment 1 Size-oriented		Experiment 2 Size-oriented		Experiment 2 Color-oriented		Experiment 3 Size-oriented	
Test type	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
Small test								
"Both" hit	.72	.04	.69	.03	.79	.02	.70	.03
Small hit	.81	.04	.80	.03	.81	.03	.81	.03
Large FA	.52	.04	.54	.03	.61	.04	.48	.04
New FA	.10	.02	.20	.04	.16	.03	.13	.03
Large test								
"Both" hit	.65	.04	.65	.03	.73	.02	.70	.03
Small hit	.83	.04	.80	.02	.82	.03	.83	.03
Large FA	.40	.04	.41	.03	.59	.04	.38	.03
New FA	.09	.02	.14	.03	.14	.03	.06	.02

Note. *SEM* = standard error of the means; Small = small-font words; Large = large-font words; "Both" hit = words in both large and small fonts; FA = false alarms.

orability illusion reported by Rhodes and Castel (2008) and extend it to several other measures of memory. Critically, as predicted by the memorial-expectation hypothesis, illusory expectations did track false recognition on the criterial recollection tests. Participants were less likely to commit false recognition when retrieval was oriented toward large words compared with small words, leading to a paradoxical increase in d' independent of actual memory effects. To our knowledge, these findings are the first to show that memorability illusions can reduce false recognition, and they provide strong evidence that memorial expectations can affect retrieval monitoring independent of memory quality or feature overlap.

Experiment 2

The results from Experiment 1 were consistent with (and predicted by) the memorial-expectation hypothesis. This interpretation assumes that there were no actual memory differences between the large and small words, but rather, participants adopted more conservative response criteria on the large test owing to illusory memorability expectations. As an alternative explanation, it could be argued that the same perceptual fluency effects that potentially drove the JOL differences at encoding (Rhodes & Castel, 2008) also influenced our test results. If large words were processed more fluently than small words at encoding, then these fluency effects may have influenced false recognition of these words at test (cf. Westerman et al., 2003; Whittlesea, 1993). Although we presented words in a neutral font size on each test in Experiment 1, words studied in large font were used as studied lures on the small test (and vice versa). If words studied in large font elicited greater fluency than did those studied in small font, then these potential fluency differences could have affected false recognition across the tests instead of memorial expectations.

Experiment 2 provided an additional test of the idea that memorial expectations were critical for the false recognition effects observed in Experiment 1, as opposed to fluency differences between the types of words across the tests. The design was similar to that in Experiment 1, but we added a color manipulation during encoding so that we could further manipulate memorial expectations at test. In addition to presenting words in large and small fonts, we experimentally yoked font size with font color such that large words were presented in red font and small words were presented in green font. This color manipulation allowed us to keep all of the test words the same but altered which features participants had to search in memory. One group of participants

Table 3
Mean d′ Scores on Each Criterial Recollection Test in Experiments 1–3

	Experiment 1 Size-oriented		Experiment 2 Size-oriented		Experiment 2 Color-oriented		Experiment 3 Size-oriented	
Test type	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
d'-studied								
d'-large	1.39	0.18	1.09	0.11	0.72	0.95	1.40	0.15
d'-small	0.91	0.14	0.81	0.13	0.67	0.14	1.03	0.14
d'-nonstudied								
d'-large	2.46	0.19	2.11	0.16	2.17	0.17	2.65	0.16
d'-small	2.32	0.17	1.95	0.20	2.11	0.18	2.26	0.17

Note. *SEM* = standard error of the means; Large and small refer to d' scores on the large-font test and small-font test, respectively.

Table 4
Mean Proportion Recognized and Subjective Judgments on the Standard Recognition Memory Test for Experiments 1 and 2

Judgment type	Experiment 1		Experiment 2			
	Size-oriented		Size-oriented		Color-oriented	
	M	*SEM*	*M*	*SEM*	*M*	*SEM*
			Old responses			
"Both" hit	.90	.02	.88	.03	.88	.03
Large hit	.91	.02	.87	.02	.88	.03
Small hit	.90	.03	.83	.03	.89	.03
New FA	.09	.02	.11	.02	.12	.02
			Actually Recollect/familiarity estimates (IRK)			
"Both" hit	.60/.83	.05/.03	.53/.73	.03/.04	.60/.75	.05/.05
Large hit	.61/.79	.05/.04	.56/.65	.03/.04	.64/.74	.05/.04
Small hit	.58/.83	.05/.03	.53/.69	.03/.04	.62/.79	.05/.04
New FA	.01/.11	.00/.02	.02/.11	.01/.01	.02/.14	.01/.02
			Amount of Unique Recollected Details			
"Both" hit	4.07	0.19	4.26	0.20	4.14	0.21
Large hit	4.05	0.18	4.47	0.16	4.31	0.19
Small hit	3.99	0.20	4.27	0.21	4.13	0.19
			d' scores			
d'-large	2.69	0.13	2.41	0.11	2.47	0.15
d'-small	2.69	0.14	2.35	0.14	2.56	0.15

Note. *SEM* = standard error of the means; Large = large-font words; Small = small-font words; "Both" hit = words in both large and small fonts; FA = false alarm; IRK = independent remember/know.

was oriented toward font size as in Experiment 1 (i.e., they took a large test and a small test), whereas another group of participants was oriented toward font color (i.e., they took a red test and a green test).

If illusory expectations relevant to font size drove the false recognition differences in Experiment 1, then this effect should be replicated when retrieval is oriented toward size but not when retrieval is oriented toward color. Although the same types of words were tested across these conditions, only retrieval orientation toward size should elicit false expectations based on the size illusion. Alternatively, if fluency effects inherent to the types of test words caused the false recognition results in Experiment 1, then we would expect the false recognition differences to persist independent of retrieval orientation. As in Experiment 1, we kept the types of words constant across conditions and changed only the retrieval orientation (via test instructions).

Method

Participants. Eighty-one students (M_{age} = 19.76, SD = 1.49) at the University of Chicago participated in the experiment for course credit. Data from one participant were removed because the participant failed to complete the experiment, resulting in 80 participants.

Materials, design, and procedure. The materials, design, and procedure were similar to those in Experiment 1 with a few exceptions. First, a between-subjects variable was added: retrieval orientation (font size, font color). Second, font size and color were yoked such that large-only words were presented in red font and small-only words were presented in green font. To minimize the ability of participants to make a strong association between font size and color, "both" words were presented in colors opposite to those for large-only and small-only words. In other words, whereas large-only words were presented in red font twice, "both" words were presented in large/green font once and small/red font a second time (and vice versa for small words). Font color was counterbalanced across participants. Third, while the size-oriented group received the same test instructions as in Experiment 1 (e.g., "Did you study this word in large font?"), the color-oriented group received test instructions oriented toward font color (e.g., "Did you study this word in red font?").

Results and Discussion

Study ratings. The pattern of study ratings for each group replicated Experiment 1 (see Table 1). The table refers to the word types as large and small even though large words also were presented in red font and small words also were presented in green font. Participants' JOLs were higher for large compared with small words and were higher when words were seen a second time compared with the first time. Also as in Experiment 1, the difference between large and small words was larger the first time the words were seen. JOLs were entered into a 2 (presentation: first, second) × 2 (size: large, small) × 2 (group: size-oriented, color-oriented) ANOVA with the first two factors as within-subject factors and the last as a between-subjects factor confirming a main effect of size, $F(1, 78) = 57.31$, $MSE = 0.85$, $p < .001$, $\eta_p^2 = .42$; a marginal effect of presentation, $F(1, 78) = 3.91$, $MSE = 0.75$, $p = .051$, $\eta_p^2 = .048$; and a Presentation × Size interaction, F(1,

78) = 13.04, $MSE = 0.075$, $p = .001$, $\eta_p^2 = .14$. No other main effects or interactions were significant.

Criterial recollection test. Results from the large and small tests can be found in Table 2. Note that Table 2 refers to large and small word types for the color-oriented group, but large words also were presented in red font and small words also were presented in green font (with the exception of "both" words, which were presented once in each font size and color; see the Method section).

The critical comparison was the level of false recognition to the studied lures across the two tests in the size-oriented group compared with the color-oriented group. Replicating Experiment 1, false recognition to studied lures on the large test was lower than false recognition on the small test in the size-oriented group, $t(39) = 3.20$, $SEM = .042$, $p = .003$, $d = 0.65$. Analogous false recognition differences also were found for nonstudied lures, $t(39) = 2.51$, $SEM = .026$, $p = .016$, $d = 0.29$. We also replicated the discrimination differences found in Experiment 1: d'-studied was greater on the large than small test, $t(39) = 2.30$, $SEM = .12$, $p = .027$, $d = 0.36$, but no significant difference was found for d'-nonstudied ($p = .22$).

Critically, participants in the color-oriented group did not show any test differences for false recognition to studied or to nonstudied lures (all $ps > .46$). Similarly, no differences were found for d'-studied tests ($p = .75$) or d'-nonstudied tests ($p = .62$; see Table 3). These results cannot be explained by differences in processing fluency between the type of words or by any other type of memory difference inherent to the words themselves, because the same types of words were tested in each case. Instead, they provide additional evidence for the impact of retrieval orientation (and corresponding illusory expectations) on memory decisions.

We also analyzed confidence judgments for each correct response between the large and small tests for each group. In the size-oriented group, no differences were found between words on the large test (hits to "both" targets $M = 5.50$, $SE = 0.14$; hits to criterial targets $M = 5.87$, $SE = 0.14$; correct rejections to studied lures $M = 4.44$, $SE = 0.17$; correct rejections to nonstudied lures $M = 5.49$, $SD = 0.17$) and the small test (hits to "both" targets $M = 5.41$, $SE = 0.16$; hits to criterial targets $M = 5.74$, $SE = 0.16$; correct rejections to studied lures $M = 4.45$, $SE = 0.19$; correct rejections to nonstudied lures $M = 5.62$, $SD = 0.15$; all $ps > .22$). In the color-oriented group, only differences in hits to criterial targets were found such that red (large) words ($M = 5.46$, $SE = 0.15$) had higher confidence ratings than did green (small) words ($M = 5.09$, $SE = 0.15$), $t(39) = 3.16$, $SEM = .12$, $p = .003$, $d = 0.39$. There were no differences in confidence ratings for hits to "both" targets ($M = 4.93$, $SE = 0.15$, and $M = 5.04$, $SE = 0.15$, for large and small tests, respectively), correct rejections to studied lures ($M = 4.11$, $SE = 0.17$, and $M = 3.97$, $SE = 0.16$, for large and small tests, respectively), or correct rejections to nonstudied lures ($M = 5.33$, $SE = 0.16$, and $M = 5.37$, $SE = 0.15$, for large and small tests, respectively; all $ps > .38$).

Standard recognition memory test. As in Experiment 1, we found no differences between large and small words on any of the measures on the standard recognition memory test (see Table 4). On the *Old/New* recognition portion of the test, hits to large targets did not differ from hits to small targets in the size-oriented group ($p = .07$) or the color-oriented group ($p = .61$). There were also no differences between large and small words for *Actually Recollect* judgments (all $ps > .37$), *Very Familiar* judgments (all $ps > .24$), familiarity estimates (all $ps > .25$), *Amount of Unique Recollected Details* (all $ps > .15$), or d' (all $ps > .24$) in each group. These results support the idea that no differences in memory strength or quality existed between large and small words in memory.

Experiment 3

The results of the preceding experiments indicated that false recognition was reduced when participants based their memory decisions on large font compared with small font. We suggest that these false recognition differences were due to the use of different response criteria across the two tests, based on the illusory expectation that large words would be remembered better than small words. Critically, this interpretation assumes that there were no actual memory differences between large and small words, and the results from our standard recognition memory test and corresponding subjective judgments were consistent with this assumption. However, one final memory-based explanation of our results needs to be addressed, one that is based on a different set of assumptions about how to interpret measures of memory accuracy (such as d') on the different tasks that we used.

According to this memory-based account, although we failed to find memory differences on the standard recognition memory tests, the false recognition differences that we obtained on the criterial recollection tests might themselves be interpreted as a kind of memory difference between large and small words. This alternative account is based on a relatively simple interpretation of signal detection theory, which is that any differences in d' must necessarily be due to an actual memory difference for the corresponding targets (assuming that memory strength for the lures is held constant). By this argument, the standard recognition tests and subjective judgments may have been relatively less sensitive to actual memory differences between large and small words compared with the criterial recollection tests, which more explicitly required participants to retrieve specific information. If large words really were stronger in memory than small words (i.e., more distinctive), then false recognition might have been lower on the large test based on the same feature-overlap process that we described in the introduction, as opposed (or in addition) to different memorial expectations.

Some evidence against this alternative memory-based account already has been presented in that our color manipulation showed no memory differences even though the same items were tested. Nevertheless, it could be argued that (a) participants must be oriented specifically toward font size features rather than other types of features in the stimuli to detect corresponding memory differences and (b) only under this retrieval orientation would false recognition be affected by resulting differences in feature overlap between the lures and the target set. While this account may seem relatively far-fetched, it remains the major alternative of our false recognition results compared with an account that is based entirely on illusory memorability expectations.

To provide a final test between these two accounts, in Experiment 3 we used a two-alternative-forced-choice (2AFC) test along with our retrieval orientation manipulation (cf. Marsh & Hicks, 1998). Using 2AFC tests is a simple way to measure memory differences between two classes of items. These tests require that participants choose between two potential targets without being able to indicate that an item was "nonstudied" (or new), thus eliminating the *Yes/No* bias inherent to other kinds of recognition memory tests (e.g., Macmillan & Creelman, 1990), including the criterial recollection task. In other words, 2AFC tests preclude the ability to change one's *Yes/No* response criteria as a function of memorial expectations. This aspect is critical because, according to the memorial-expectation account, the false recognition effects we found in the prior experiments were due to more conservative response criteria when retrieval was oriented toward large font compared with small font. These expectations cannot affect 2AFC tests because these tests do not have a nonstudied option. Thus, the memorial-expectation account predicts no effects of retrieval orientation on 2AFC test accuracy, because it assumes no differences in memory exist between large and small words. In contrast, the memory-based account does predict an effect of retrieval orientation on the 2AFC test. According to this account, large words are more memorable than small words, and orienting retrieval toward large font information capitalizes on these differences in memory. This alternative account might explain the accuracy differences we found across the different retrieval orientations on the criterial recollection test, and based on the same reasoning, it predicts accuracy differences across different retrieval orientations on a 2AFC test.

As in Experiment 1, participants studied large and small words and then were tested on neutral-sized words under different retrieval orientation instructions. However, on each test participants were presented with word pairs for a 2AFC judgment instead of a *Yes/No* judgment. One 2AFC test required participants to select the item in the test pair that had been studied as a large word, whereas the other 2AFC test required them to select the item in the test pair that had been studied as a small word. For the critical test pairs, one word in the pair was previously presented in the criterial format (e.g., large words on the large test), and the other word was previously presented in the noncriterial format (e.g., small words on the large test). Thus, this test required the recollection of size information to distinguish between the words in the test pair (chance = 50%). We also included noncritical test pairs, in which the target item had been studied in both sizes. Similar to the prior experiments, these "both" targets were included so that the recollection of the noncriterial format for one of the words in the test pair would not preclude that word as having also been presented in the criterial format. Thus, even though the same critical test pairs occurred on each test block, participants needed to selectively base these 2AFC decisions on memory for large words on the large test and small words on the small test.

If large words are more memorable than small words, and critically, if these features can be differentially retrieved based on our retrieval orientation instructions, then participants should show differences in memory accuracy on the 2AFC tests (large test > small test). This outcome is predicted by the memory-based account of our prior results in the criterial recollection task. In contrast, if large and small words are equally memorable, then participants taking the 2AFC tests should not show memory accuracy differences depending on whether they are oriented toward large or small font at test. This outcome would bolster the hypothesis that our previous criterial recollection results were due to illusory memorial expectations, as opposed to actual differences in memory. We also had an independent group of participants take criterial recollection tests similar to those in Experiments 1 and 2 in an attempt to replicate the key false recognition results one final time.

Method

Participants. Sixty-six students (M_{age} = 20.14, *SD* = 1.57) at the University of Chicago participated in the experiment for course credit or $10. Data from two participants were discarded (one was given the wrong script, and one started the tests before being given instructions), resulting in 64 participants.

Materials and design. The current experiment consisted of a 2 (font size: large, small) × 2 (test orientation: large, small) × 2 (test type: criterial recollection, two-alternative-forced-choice) design with font size and test orientation as within-subject factors and test type as a between-subjects factor. During the study phase, participants were presented with 144 words taken from the same set as in the previous experiments. Because of the different nature of each test, participants received different proportions of each item type. Participants receiving the criterial recollection tests viewed 48 words in large font only, 48 in small font only, and 48 in both large and small font ("both" words), and they never studied an additional set of 48 words. Words were counterbalanced across the four types (large-only, small-only, "both," and new). Participants receiving the 2AFC tests viewed 54 words in large font only, 54 in small font only, and 36 in both large and small fonts. Words were counterbalanced across the three types of words. No additional words were set aside to serve as nonstudied items on the 2AFC tests. As in the prior experiments, we repeated the large-only and small-only words so that each would be presented twice, thereby equating their number of presentations with the items that had been studied once in each size (i.e., "both" items). In this way, the relative number of study presentations could not be used to differentiate targets from lures on the subsequent tests, so that participants instead had to recollect size information.

After the study phase, half the participants received two criterial recollection tests, and half the participants received 2AFC tests with test order and test type counterbalanced across participants. The criterial recollection tests were identical to those in Experiment 1 except that each test consisted of 96 words (randomly arranged) made of 24 of each type (large-only, small-only, "both," and new). The 2AFC tests consisted of two types of pairs: 18 target–lure pairs (critical pairs) and 18 "both" target–lure pairs (noncritical pairs). For instance, on the large test, participants either saw a pair of words that consisted of a "both" word and a small word or saw a pair of words that consisted of a large word and a small word. Half of the time, targets were presented on the left, and the other half, on the right, with order counterbalanced across participants. Each pair of words at test was presented with the appropriate test prompt (e.g., "Which word did you study in large font?" during the large test). Targets were words that corresponded to the appropriate test (e.g., large words for the large–small pairs on the large test, and "both" words for the "both"–small pairs on the large test), and studied lures were words that did not correspond to the appropriate test (e.g., small words for the large–small pairs and also for the "both"–small pairs on the large test).

Procedure. The study phase was identical to that of Experiment 1 for participants receiving both the criterial recollection tests and the 2AFC tests. For participants receiving the criterial recollection tests, the test procedure was identical to Experiment 1 with the exception that there was no standard recognition memory test. For participants receiving the 2AFC tests, participants also took two criterial recollection tests (large and small test), which were modified for the 2AFC format. On the large test, participants were instructed to press the key labeled *LEFT* if they remembered viewing the word on the left in large font (large-only or "both" words) and to press the key labeled *RIGHT* if they remembered viewing the word on the right in large font. They were also told that remembering that a word was viewed in small font was irrelevant on this test because, for some words, they viewed the word in both large and small font. On both types of tests, participants were then asked to make confidence judgments on a scale from 1 (*Low*) to 7 (*High*). Analogous instructions were given for the small test, with the only change being in the to-be-recollected (or criterial) information.

Results and Discussion

Study ratings. The JOL ratings for both groups of participants followed a pattern that was similar to that of the prior experiments (see Table 1). Participants' JOLs were rated higher for words in large than small font, and JOLs were rated higher for words seen a second time than when they were seen the first time. JOLs were entered into a 2 (presentation: first, second) × 2 (size: large, small) × 2 (test: criterial recollection, 2AFC) repeated-measures ANOVA, revealing a main effect of presentation, $F(1, 62) = 18.88$, $MSE = 1.05$, $p < .001$, $\eta_p^2 = .23$, and a main effect of size, $F(1, 62) = 91.51$, $MSE = 0.25$, $p = .001$, $\eta_p^2 = .60$, and qualified by a Presentation × Size interaction, $F(1, 62) = 19.26$, $MSE = 0.071$, $p = .001$, $\eta_p^2 = .24$. The interaction indicated that the difference between large and small JOL ratings was larger when seen the first time than the second. No other main effects or interactions were significant.

Criterial recollection test. The criterial recollection results replicated those found in Experiment 1 and Experiment 2 (size-oriented condition; see Tables 2 and 3). False recognition to studied lures was reduced on the large test compared with the small test, $t(31) = 2.73$, $SEM = .036$, $p = .01$, $d = 0.48$. In this experiment, false recognition to nonstudied lures also was reduced on the large test compared with the small test, $t(31) = 2.07$, $SEM = .030$, $p = .046$, $d = 0.47$. However, there were no differences between hits to "both" targets ($p = .98$) or to criterial targets ($p = .33$). We also found a greater d'-studied difference for the large test than small test, $t(31) = 3.49$, $SEM = .11$, $p = .001$, $d = 0.47$, and a greater d'-nonstudied difference for the large test than small test, $t(31) = 2.62$, $SEM = .15$, $p = .013$, $d = 0.42$. Confidence ratings were greater on the large than small test for hits to "both" targets ($M = 5.32$, $SE = 0.20$, and $M = 5.05$, $SE = 0.20$, on the large and small tests, respectively), $t(31) = 2.55$, $SEM = .10$, $p = .016$, $d = 0.23$, and to criterial targets ($M = M = 5.69$, $SE = 0.18$, and $M = 5.37$, $SE = 0.22$, on the large and small tests, respectively), $t(31) = 2.60$, $SEM = .12$, $p = .014$, $d = 0.29$, similar to results in Experiments 1 and 2. Confidence ratings did not differ between tests ($ps > .14$) for correct rejections to studied lures ($M = 3.99$, $SE = 0.22$, and $M = 4.17$, $SE = 0.19$, on the large and small tests, respectively) and to nonstudied lures ($M = 5.02$, $SE = 0.28$, and $M = 5.03$, $SE = 0.24$, on the large and small tests, respectively).

Two-alternative-forced-choice test. In contrast to our results using the criterial recollection tests, there were no differences in memory accuracy for the critical target–lure pairs on the 2AFC tests ($M = 0.82$, $SE = 0.03$, and $M = 0.79$, $SE = 0.03$, for large and small tests, respectively), nor were there any differences for the noncritical "both" target–lure pairs ($M = 0.72$, $SE = 0.03$, and $M = 0.70$, $SE = 0.03$, for large and small tests, respectively) on the large test compared with the small test (all $ps > .29$). This outcome is inconsistent with the memory-based account that retrieval orientation enhances memory differences between large and small words but was predicted by the memorial-expectation account. Confidence ratings for correct responses to target–lure pairs on the large test ($M = 5.30$, $SE = 0.20$) were greater than on the small test ($M = 4.96$, $SE = 0.21$), $t(31) = 2.80$, $SEM = .12$, $p = .009$, $d = 0.29$, again consistent with the prior experiments. However, there was no confidence difference for "both" target–lure pairs ($M = 4.80$, $SE = 0.19$, and $M = 4.71$, $SE = 0.19$, on the large and small tests, respectively; $p = .50$).

In sum, participants reduced false recognition on the large compared with the small test when given criterial recollection tests, consistent with the idea that memorial expectations exert an independent effect on false recognition. These results provide a second replication of this illusory expectation effect. However, when given a 2AFC test, which prevented the use of memorial expectations to suppress false recognition (i.e., preventing the use of more conservative recollection-based criteria to make a *Yes/No* decision), no memory accuracy differences were found between large and small words. The 2AFC results suggest that features recollected when oriented toward large font were not better remembered than were features recollected when oriented toward small font, thereby undermining a key assumption of the memory-based account of our criterial recollection results. This test provides additional evidence that large words are not better remembered than small words, again suggesting that the differences in false recognition on the criterial recollection task were not due to actual differences in memory.

General Discussion

The present work was aimed at understanding how illusory expectations can shape memory decisions. Because memory is fallible, we often have to monitor the information we retrieve to decide what occurred in the past. It has long been implied that we hold relatively accurate expectations about different sources of memories and that these expectations can guide monitoring decisions and affect accuracy and distortion (Johnson et al., 1993). The novel finding from the present study is that illusory expectations also can influence memory accuracy, beyond any real memory differences. These experiments provide some of the strongest evidence that expectations can suppress false recognition independent of actual differences in memory, demonstrating the independent role of memorial expectations in the retrieval-monitoring process.

The current study was designed to isolate expectations from actual memory differences, which have historically been difficult to disentangle. To this end, we took advantage of a natural memorability illusion, rather than misleading participants or otherwise attempting to artificially create response biases. In Experiment 1, we tested whether the false belief that large words were better remembered than small words could affect false recognition, independent of the degree of overlapping features in memory. As predicted, false recognition was reduced when searching memory for large compared with small words, and this difference was tracked by participants' expectations (as measured by JOL ratings). These results suggest that distinctive encoding is not necessary to influence memorial expectations, and they are consistent with previous findings that used misleading instructions at test to manipulate expectations (Dodson & Schacter, 2002; Westerman et al., 2003). In Experiment 2, we further ruled out memory-based explanations (e.g., fluency accounts) for the differences in false recognition by manipulating participants' retrieval orientation at test. When retrieval was oriented toward size information, we replicated the key false recognition differences of Experiment 1, but when participants no longer searched for features that differed in memorial expectations (i.e., font color rather than font size information), these false recognition differences were eliminated, even though participants were tested with the same types of words across the different retrieval orientations.

The memorial-expectation account of these results rests on the assumption that large words are not better remembered than small words and that this expectation is, in fact, an illusion (Rhodes & Castel, 2008). Otherwise one might argue that potential memory differences between large and small words affected false recognition via a feature-overlap process. As a check of this assumption we included standard recognition tests and subjective judgments in the first two experiments, and in Experiment 3 we used a 2AFC test, which precluded recognition effects due to shifts in *Yes/No* response bias. We found little evidence to support the idea that large words were better remembered than small words on any of our tests, as indexed by criterial recollection hits, standard *Old/New* recognition hits, 2AFC hits, *Actually Recollect* judgments, *Very Familiar* judgments, *independent remember/know* familiarity estimates, unique recollected detail judgments, or d' scores from the standard recognition test. Across experiments, these comparisons provided 19 different opportunities to measure actual memory differences between large and small words (excluding "both" words on each test), but none were statistically significant at the traditional $p < .05$ threshold, and most were not even close.

In order for a feature-overlap hypothesis (or other memory-based hypothesis) to explain the current results, several post hoc assumptions would have to be made. Specifically, the hypothesis would have to assume (a) that large words were better remembered than small words; (b) that the criterial recollection tests had greater sensitivity to detect these memory differences than had more standard measures, but only when retrieval was oriented toward font information; and (c) that for some unknown reason, these target memory differences were more likely to be reflected in false recognition differences on the criterial recollection test than by differences in the actual targets themselves.[4] We see this interpretation as less likely than an expectation-based account, given the sheer number of null results in memory for studied items that we obtained, in addition to prior research investigating memory for large and small words that found null results using other kinds of memory tests (Rhodes & Castel, 2008; see also Kolers, Duchnicky, & Sundstroem, 1985; Robinson & Standing, 1990). Rather, we interpret the false recognition differences that we obtained on the criterial recollection tests as being due to illusory memorial expectations.

The current results also cannot be entirely explained by appealing to differences in recollection and familiarity between the types of words, as in some dual process theories of the mirror effect (e.g., Cary & Reder, 2003; Joordens & Hockley, 2000). Dual process theories have been proposed to explain false recognition differences in distinctive contexts. For instance, Cary and Reder (2003) suggested that when unable to rely on recollection, people are more likely to rely on familiarity and use more liberal familiarity criteria, increasing false recognition. In contrast, when people are able to rely on recollection, they will rely less on familiarity and use more conservative familiarity criteria, decreasing false recognition. However, neither recollection nor familiarity differed between the types of words in the present experiment, so some other factor (such as illusory expectations) must have played a role in the shifting of the response criteria across our criterial recollection tests. Although we agree that recollection and familiarity are likely to make unique contributions to false recognition in other contexts, such differences cannot explain the current results. Retrieval-monitoring processes can influence response criteria and corresponding false recognition rates independent of these memory-based processes.

Although we have appealed to the concept of response criteria to explain the false recognition effects that we observed on the criterial recollection task, it is important to point out that the concept of response criteria has been employed in several different ways in the literature. Depending on one's adopted framework or assumptions about memory, a "change" or "shift" in response criteria could refer to (a) shifting one's criterion along a single familiarity dimension (as in classic signal detection theory; cf. Stretch & Wixted, 1998), (b) shifting one's reliance between familiarity and recollection (as in dual process theories; cf. Yonelinas, 2002), (c) shifting between different kinds of recollected features that one uses in the memory decision (as in the source-monitoring framework; cf. Johnson et al., 1993), or (d) some combination of all three.

Although the size illusion might have affected expectations about recollection or familiarity, there is some reason to suspect that recollection-based expectations played a key role in our false recognition effects. First, because large and small words were matched on familiarity, we assume that the ability to discriminate between these items on the criterial recollection tests was based on the recollection of size information and corresponding recollection-based response criteria. Second, a recent study by Scimeca, McDonough, and Gallo (2011) showed that a direct manipulation of familiarity (study repetitions) did not elicit false recognition differences on the criterial recollection task as one would expect if familiarity-based expectations could affect response criteria. Instead, this study identified differences in recollection as more critical for finding these kinds of false recognition effects. More specifically, differences in recollective distinctive-

[4] From a signal detection perspective, it could be that participants set their response criterion to achieve the same rate of target hits on each criterial recollection test, while allowing false recognition of lures to vary as a function of the strength of the lures (due to feature overlap or other processes). However, it is unclear to us why participants would do this or whether they even could, because at the present time, this is only a post hoc interpretation.

ness were key (i.e., expecting that one class of items would elicit qualitatively richer or more-unique memories than another). When interpreted in this way, the size illusion effects found in the current study might reflect an illusory expectation about recollective distinctiveness, although our JOLs did not differentiate between different kinds of memory in this way, so this idea awaits further testing.

One common question about the memorial-expectation hypothesis that we have described here is why the proposed changes in response criteria are not evident for all of the test items ("both" targets, criterial targets, studied lures, and nonstudied lures). Although a global shift in response criteria does entail that all test items could potentially be evaluated under the new criteria, it does not necessarily follow that the response rates to different kinds of test items will be equally affected by the new criteria. For example, within even the simplest of signal detection theories, criterion shifts could differentially affect the recognition rates to targets and lures depending on the distance between the criteria placements and the means of each relevant distribution (i.e., a shift in the tail of a distribution would have a smaller effect than a shift near the center). Perhaps even more relevant to the current studies is the idea that different kinds of test items may elicit different degrees of recollection and familiarity. As discussed in the introduction, test items that elicit the retrieval of vague or questionable evidence (i.e., a strong feeling of familiarity in the absence of a clear recollection) may be most likely to engage retrieval-monitoring processes that result in the adjustment of retrieval criteria.

Related to this last point, there is evidence that memorial expectations and subsequent retrieval-monitoring processes may play a larger role when item-specific details are not recollected or when only a general sense of familiarity is present (cf. Schacter et al., 1999). Support for this idea comes from the repetition-lag paradigm, in which nonstudied lures were repeated with various amounts of intervening items after participants studied pictures or words (Dodson & Schacter, 2002). Dodson and Schacter (2002) found that reductions in false recognition following the encoding of pictures relative to words increased as the repetition lag increased, potentially because participants were less likely to recollect item-specific information regarding the word's origin at longer lags, resulting in a greater reliance on the distinctiveness heuristic. Additional evidence using the same task comes from a study comparing healthy older adults with participants with probable Alzheimer's disease (Budson, Dodson, Daffner, & Schacter, 2005). In contrast to cognitively normal participants, in this study participants with Alzheimer's disease showed reduced recognition in the picture condition for both targets and lures. The authors attributed this additional reduction in hits when searching memory for pictures relative to words to the impoverished target recollections caused by Alzheimer's disease that did not reach the participants' expected response criteria. Thus, memorial expectation effects may be most evident under conditions in which item-specific recollections cannot be adequately retrieved, but the test items nevertheless elicit a strong sense of familiarity that requires additional monitoring.

The present study also suggests that memorial expectations can affect confidence ratings (cf. Busey, Tunnicliff, Loftus, & Loftus, 2000; Reinitz, Peria, Séguin, & Loftus, 2011). A frequent finding across the three experiments was that large hits were endorsed with greater confidence than were small hits. Although confidence judgments always lend themselves to careful scrutiny because of their subjective nature, they are sometimes the sole measure of objective memory differences. For example, they often contribute to discrimination measures in signal detection studies that use confidence judgments to construct receiver operating characteristics. Studies using these measures frequently do not take into account metacognitive factors such as memorial expectations, which may have a significant impact on both memory accuracy and confidence ratings. In fact, the current study suggests that derived accuracy statistics from signal detection theory (such as d') can be affected by processes other than memory quality, such as illusory expectations, that disproportionately affect false recognition. While these kinds of factors may not be problematic in most typical applications of the d' measure, the current results suggest that they may need to be more carefully considered.

In conclusion, the finding that memorial expectations can influence false recognition independent of actual memory differences has broad implications. Recent work has suggested that expectations may especially impact memory performance for vulnerable populations such as healthy older adults (e.g., Hertzog & Hultsch, 2000; McDaniel, Einstein, & Jacoby, 2008), patients with brain damage (e.g., Kit, Tuokko, & Mateer, 2008; Scholl & Sabat, 2008), or people who believe they have an altered mind-state via placebos (e.g., Assefi & Garry, 2003). In healthy older adults, for example, there is mounting evidence that expectations of cognitive decline via natural aging can induce stereotype threat. In addition to impairing general cognitive capacities such as working memory (e.g., Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003), stereotype threat may further influence older adults by making them more reluctant to use memory at all (Touron & Hertzog, 2004). While much prescriptive research has focused on how to improve retrieval access or the quality of memories themselves, less work has focused on altering expectations in the memory literature, which can exert an independent influence on memory accuracy. Although the current study was not designed to address these applied issues, it does demonstrate the central importance of memorial expectations in the retrieval-monitoring process.

References

Assefi, S. L., & Garry, M. (2003). Absolut® memory distortions: Alcohol placebos influence the misinformation effect. *Psychological Science, 14,* 77–80. doi:10.1111/1467-9280.01422

Bartlett, F. C. (1932). *Remembering.* Oxford, England: University Press.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136,* 256–276. doi:10.1037/0096-3445.136.2.256

Budson, A. E., Dodson, C. S., Daffner, K. R., & Schacter, D. L. (2005). Metacognition and false recognition in Alzheimer's disease: Further exploration of the distinctiveness heuristic. *Neuropsychology, 19,* 253–258. doi: 10.1037/0894-4105.19.2.253

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7,* 26–48. doi:10.3758/BF03210724

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248. doi:10.1016/S0749-596X(03)00061-5

Chappell, M., & McClelland, J. L. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760. doi:10.1037/0033-295X.105.4.734-760

Cleary, A. M., Morris, A. L., & Langley, M. M. (2007). Recognition memory for novel stimuli: The structural regularity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 379–393. doi:10.1037/0278-7393.33.2.379

Criss, A. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55,* 461–478. doi:10.1016/j.jml.2006.08.003

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46,* 782–803. doi:10.1006/jmla.2001.2822

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38,* 833–848. doi:10.3758/MC.38.7.833

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of item distinctiveness. *Psychology and Aging, 22,* 209–213. doi:10.1037/0882-7974.22.1.209

Gallo, D. A., Meadow, G. N., Johnson, E. L., & Foster, K. T. (2008). Levels of processing and distinctiveness: Evidence from the criterial recollection task. *Journal of Memory and Language, 58,* 1095–1111. doi:10.1016/j.jml.2007.12.001

Gallo, D. A., Perlmutter, D. H., Moore, C. D., & Schacter, D. L. (2008). Distinctive encoding reduces the Jacoby-Whitehouse illusion. *Memory & Cognition, 36,* 461–466. doi:10.3758/MC.36.2.461

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473–493. doi:10.1016/j.jml.2004.06.002

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19,* 617–623. doi:10.3758/BF03197157

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action items: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26,* 20–33. doi:10.3758/BF03211367

Greene, R. L. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50,* 259–267. doi:10.1016/j.jml.2003.12.001

Greene, R. L. (2007). Foxes, hedgehogs, and mirror effects: The role of general principles in memory research. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 53–66) New York, NY: Psychology Press.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127,* 251–268. doi:10.1037/0096-3445.127.3.251

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed.; pp. 417–466) Mahwah, NJ: Erlbaum.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126–135. doi:10.1037/0096-3445.118.2.126

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi:10.1037/0033-2909.114.1.3

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85. doi:10.1037/0033-295X.88.1.67

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1534–1555. doi:10.1037/0278-7393.26.6.1534

Kit, K. A., Tuokko, H. A., & Mateer, C. A. (2008). A review of the stereotype threat literature and its application in a neurological population. *Neuropsychology Review, 18,* 132–148. doi:10.1007/s11065-008-9059-9

Kolers, P. A., Duchnicky, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 726–751. doi:10.1037/0096-1523.11.6.726

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 954–963. doi:10.1037/0278-7393.31.5.954

Lyle, K. B., Bloise, S. M., & Johnson, M. K. (2006). Age-related binding deficits and the content of false memories. *Psychology and Aging, 21,* 86–95. doi:10.1037/0882-7974.21.1.86

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413. doi:10.1037/0033-2909.107.3.401

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1137–1151. doi:10.1037/0278-7393.24.5.1137

McDaniel, M. A., Einstein, G. O., & Jacoby, L. L. (2008). New considerations in aging and memory: The glass may be half full. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed.; pp. 251–310) New York, NY: Psychology Press.

McDonough, I. M., & Gallo, D. A. (2008). Autobiographical elaboration reduces memory distortion: Cognitive operations and the distinctiveness heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1430–1445. doi:10.1037/a0013013

McDonough, I. M., & Gallo, D. A. (2010). Separating past and future autobiographical items in memory: Evidence for a reality monitoring asymmetry. *Memory & Cognition, 38,* 3–12. doi:10.3758/MC.38.1.3

Moscovitch, M., & Craik, F. I. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 447–458. doi:10.1016/S0022-5371(76)90040-2

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46) New York, NY: Oxford University Press.

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 485–497. doi:10.1037/0278-7393.3.5.485

Pierce, B. H., & Gallo, D. A. (2011). Encoding modality can affect memory accuracy via retrieval orientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 516–521. doi:10.1037/a0022217

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102. doi:10.3758/BF03211168

Reinitz, M. T., Peria, W. J., Séguin, J. A., & Loftus, G. R. (2011). Different confidence-accuracy relationships for feature-based and familiarity-based memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 507–515. doi:10.1037/a0021961

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615–625. doi:10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16,* 550–554. doi:10.3758/PBR.16.3.550

Robinson, K. M., & Standing, L. (1990). Effects of size changes on memory for words and pictures. *Perceptual and Motor Skills, 71,* 919–922. doi:10.2466/PMS.71.7.919-922

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100. doi:10.1006/jmla.1996.0005

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4,* 108–115. doi:10.1016/S1364-6613(00)01445-5

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24. doi:10.1006/jmla.1998.2611

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289–318. doi:10.1146/annurev.psych.49.1.289

Schacter, D. L., & Wiseman, A. L. (2006). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 89–107) New York, NY: Oxford University Press.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440–452. doi:10.1037/0022-3514.85.3.440

Scholl, J. M., & Sabat, S. R. (2008). Stereotypes, stereotype threat and ageing: Implications for the understanding and treatment of people with Alzheimer's disease. *Ageing & Society, 28,* 103–130. doi:10.1017/S0144686X07006241

Scimeca, J. M., McDonough, I. M., & Gallo, D. A. (2011). Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language, 65,* 363–377.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5,* 710–715. doi:10.3758/BF03208850

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50. doi:10.1037/0096-3445.117.1.34

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396. doi:10.1037/0278-7393.24.6.1379

Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition, 32,* 298–310. doi:10.3758/BF03196860

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385–398. doi:10.1037/0003-066X.40.4.385

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373. doi:10.1037/h0020071

Verde, M. F., Stone, L. K., Hatch, H. S., & Schnall, S. (2010). Distinguishing between attributional and mnemonic sources of familiarity: The case of positive emotion bias. *Memory & Cognition, 38,* 142–153. doi:10.3758/MC.38.2.142

Watkins, M. J. (1991). An experimental psychologist's view of cognitive science. In R. G. Lister & H. J. Weingartner (Eds.), *Perspectives on cognitive neuroscience* (pp. 132–144). New York, NY: Oxford University Press.

Westerman, D. L., Miller, J. K., & Lloyd, M. E. (2003). Change in perceptual form attenuates the use of the fluency heuristic in recognition. *Memory & Cognition, 31,* 619–629. doi:10.3758/BF03196102

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235–1253. doi:10.1037/0278-7393.19.6.1235

Winograd, E., Karchmer, M. A., & Russell, I. S. (1971). Role of encoding unitization in cued recognition memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 199–206. doi:10.1016/S0022-5371(71)80014-2

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. doi:10.1006/jmla.2002.2864

Received January 14, 2011
Revision received July 15, 2011
Accepted August 12, 2011 ■