

The role of recollection and familiarity in the context variability mirror effect

GABRIEL I. COOK and RICHARD L. MARSH
University of Georgia, Athens, Georgia

and

JASON L. HICKS
Louisiana State University, Baton Rouge, Louisiana

Context variability refers to the number of preexperimental contexts that are associated with concepts. In four experiments, we investigated the basis for increased recognition memory for low context variability words. Low context variability was associated with greater recollection in the hit rates, and high context variability was associated with greater familiarity in the false alarms. Shortening the study time reduced recollection, but low context variability still influenced recollection in the hit rates. A modality change from study to test also reduced recollection but preserved recollective differences for low versus high context variability items. One interpretation of the results suggests that low context variability evokes more specific and, perhaps, idiosyncratic recollective associations during learning and that these associations support better recognition in the hit rates. By contrast, activating the larger number of associations for high context variability items may be mistaken for familiarity in the false alarm rates.

A mirror effect occurs in recognition memory tests when a particular class of items produces a higher hit rate and a lower false alarm rate, as compared with another class of items. For example, low-frequency words usually yield a greater hit rate than do high-frequency words, but false alarms are made less often to low-frequency words than to high-frequency words (see, e.g., Glanzer & Adams, 1985; Hirshman & Arndt, 1997; Hockley, 1994). A variety of stimulus comparisons produce mirror effects, such as manipulations of meaningfulness, frequency, concreteness, pictures versus words, and so forth. Steyvers and Malmberg (2003) added to the list of variables producing a mirror effect by reporting that hit rates were mirrored in false alarm rates on the basis of a dichotomy of high versus low normative context variability. They defined *context variability* as the number of preexperimental contexts in which a word is encountered (cf. Dennis & Humphreys, 2001). A word such as *summit* is a low context variability word, because it is encountered mainly in conversations and texts having to do with earth formations or mountain climbing. By contrast, a word such as *meal* is encountered in many more preexperimental contexts in everyday life and, therefore, is a high context variability word. As with word frequency, low context variability yielded higher hit rates and lower false alarm rates, as compared with high context variability items. The goal of the present

study was to explore the basis of the context variability effect. To do so, we used the remember-know procedure to obtain approximations of recollection and familiarity, in order to understand which component of recognition contributes to the hit rate and false alarm rate advantages (e.g., Jacoby, Yonelinas, & Jennings, 1997; Kishiyama & Yonelinas, 2003).

One working hypothesis is that the context variability mirror effect is similar in kind to the word frequency effect. If this is so, context variability should behave in a manner consistent with, say, Joordens and Hockley's (2000) account of the word frequency mirror effect. They argued elegantly, from both data and theory, that the hit rate advantage enjoyed by low-frequency words is due to greater recollection from the study episode (cf. Guttentag & Carroll, 1997; Reder et al., 2000). When recollection is reduced by longer retention intervals or through very brief encoding, the hit rate advantage for low-frequency words is attenuated or eliminated, but the word frequency differences in the false alarm rates remain unchanged (see also Malmberg & Nelson, 2003). According to their account, preexperimental familiarity drives the differences in the false alarm rates, and therefore, high-frequency words will virtually always yield higher false alarm rates than will low-frequency words. Joordens and Hockley, as well as Gardiner and Java (1990), have found greater proportions of remember (R) responses for low- than for high-frequency words for hits. For false alarms, the raw proportion of know (K) responses is usually greater for high-frequency than for low-frequency words but is occasionally equivalent for both classes of items. Reder and her colleagues have agreed, in spirit anyway, with the con-

We thank Christopher Berardi and Gene Brewer for their valuable assistance in collecting the data. Correspondence concerning this article should be addressed to R. L. Marsh, Department of Psychology, University of Georgia, Athens, GA 30602-3013 (e-mail: rlmars@uga.edu).

clusions drawn by Joordens and Hockley. For example, Cary and Reder (2003) extended the greater recollection in hits and greater familiarity in false alarms to list length and encoding strength effects. Arndt and Reder (2002) also demonstrated that receiver-operating characteristic analyses confirm this approach.¹

If context variability behaves like word frequency, the hit rate advantage for low context variability words should accrue from more recollection, as measured by a greater proportion of R responses, for these items than for high context variability items. By contrast, K responses could be either more frequent for high context variability items or equivalent across the two classes of items. Which of these two alternatives will occur depends on whether low context variability contributes additional recollection or whether these items trade familiarity for recollection. In the false alarms, more should occur for high than for low context variability items, and more should be labeled as K responses. As the reader will see, we controlled for word frequency, so the predicted greater proportion of K responses for high context variability items cannot be attributed to preexperimental familiarity *qua* frequency. From this perspective, there may actually be no basis for predicting more K responses in false alarms, except that R responses to new items are relatively uncommon (for a good summary, see Gardiner & Richardson-Klavehn, 2000).

More generally, our working hypothesis that context variability acts like word frequency may be incorrect, and these predictions are not necessarily preordained. Accurate predictions hang in the balance of the underlying cause for both the word frequency and the context variability effects. In the case of the word frequency effect, no agreement exists on why it occurs. The predictions made in the previous paragraph implicitly assume the multiple retrieval processes account advocated by Joordens and Hockley (2000). Other accounts of the word frequency effect include differences in the distribution of attention that favors low word frequency (e.g., Glanzer & Adams, 1990) or differences in encoding variability (e.g., McClelland & Chappell, 1998). In the case of context variability, Steyvers and Malmberg (2003) did not specify a locus for the effect. But one could easily predict that the low context variability items enjoy a larger increment to undifferentiated familiarity after being studied. Increased allocation of attention toward low context variability items or the activation of more features in these items need not result in more recollection but, rather, only in an increment to their familiarity. Jacoby (1999), as well as Gardiner and Richardson-Klavehn (2000), discussed variables such as repetition and amount of maintenance rehearsal at study that increase only familiarity. This scenario would play out in the proportion of K responses (rather than R responses), with those responses being greater for low context variability hits and greater for high context variability false alarms.

Yet another alternative outcome is that low context variability items may enjoy both more recollection and more familiarity, as was recently found for novel items in a recognition study list (Kishiyama & Yonelinas, 2003).

In this case, both R and K responses would be increased for hits, but greater false alarms to high context variability items should still be localized only in K responses. Neither theory nor past findings would allow us to unambiguously specify the predicted contributions of recollection and familiarity when context variability is manipulated. And this was exactly the reason for conducting this present study—namely, to investigate the underlying recognition processes that may account for the context variability mirror effect.

Dennis and Humphreys (2001), as well as Steyvers and Malmberg (2003), argued that context variability is naturally quite highly correlated with word frequency. High-frequency words tend to appear in many preexperimental contexts, and low-frequency words have less opportunity to appear in a variety of contexts. Nevertheless, the word frequency variable must be controlled and unconfounded with context variability, in order to isolate precisely the different contributions of recollection and familiarity at different levels of context variability. One approach would be to hold word frequency constant and simply make comparisons across low and high levels of context variability. Another approach would be to adopt Steyvers and Malmberg's strategy of orthogonally manipulating both variables and examine the influence of both word frequency and context variability simultaneously. Because it has the advantage of providing a touchstone to past work in which word frequency has been manipulated, the orthogonal manipulation of both variables was the approach adopted here.

One final preliminary issue concerns our use of the remember-know procedure. How R and K responses are made and what they represent are issues that have been debated over the last decade. We do not intend in this article to weigh in on these debates, and our study was not designed to do so. Our approach is merely to use R responses as a proxy measure of recollection, in the same way that these responses often are used to closely index estimates of recollection from a process dissociation procedure (e.g., Jacoby et al., 1997; Kelley & Jacoby, 1998). This use of R responses is less debated than is the question of how K responses map onto familiarity. Joordens and Hockley (2000) reported raw proportions of K responses, which we will do as well. However, we have adopted the approach, advocated by Yonelinas and Jacoby (1995), that K responses must be transformed before they can represent an accurate estimate of the underlying contribution of familiarity. To do so requires that we make an assumption about the relationship between recollection and familiarity, and we have chosen to assume independence, because none of the conditions shown to violate the independence assumption appears to be operative in the present study (see Jacoby, 1998). Consequently, in addition to R and K responses, we also report corrected K responses [$K/(1-R)$] as our estimate of the contribution of familiarity to recognition. Our approach specifically ignores one recent report that showed conditions in which the underlying processes ostensibly tapped by the remember-know

procedure were correlated (Humphreys et al., 2003). We acknowledge that no measure of memory is perfect, and the remember-know procedure has its critics. Nevertheless, we believe that when used correctly with detailed instructions to participants, it yields estimates that track movements in the manipulation of independent variables such as context variability.

By way of overview, we conducted four experiments with the same basic design. Words were studied that came from one of four classes created by orthogonally crossing low versus high context variability with low versus high word frequency. After learning how to make R and K responses, the participants took a recognition test with *remember*, *know*, and *new* as the three response options (see Hicks & Marsh, 1999). Experiment 1 served as an initial examination of the relative contributions of recollection and familiarity that result from manipulations of context variability. In the remaining experiments, we examined how context variability behaved in response to manipulations that were known to affect the word frequency effect or that were predicted to dissociate the basis of recognizing high and low context variability items. As such, Experiments 2 and 3 manipulated study time, and Experiment 4 manipulated study-test modality compatibility.

EXPERIMENT 1

Method

Participants. Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each of the 31 participants was tested individually in a session that lasted approximately 25 min.

Materials and Procedure. We selected 268 items from the Steyvers and Malmberg (2003) list, with 67 items in each of the four cells created by orthogonally crossing high and low word frequency with high and low context variability. Only 6 items were removed from each of the four item classes, because we thought that they might be offensive to some participants (e.g., *urine*), and so, we assume that the present materials have functionally the same characteristics as those reported by Steyvers and Malmberg. For each participant tested, the software randomly selected anew 20 items from the 67 possible items in each of the four item classes. Therefore, the participants studied a total of 80 items in a completely random order, without regard to item class. The words appeared in the center of the computer screen for 2 sec and were accompanied by a short orienting beep. In the study list, each subsequent item replaced the previous one. The participants had been instructed to pay attention to the words in preparation for an unspecified memory test. The test list was constructed by rerandomizing the items in the study list, along with an additional 80 distractor items, for a total of 160 test trials. The 80 distractor items consisted of 20 items randomly sampled from the 47 remaining items in each of the four item classes.

When the study phase was over, the participants spent several minutes reading detailed instructions on how to make R versus K judgments. These instructions were identical to the ones that we had used previously (e.g., Hicks & Marsh, 1999). After reading the instructions, the experimenter verbally reiterated them and verified that the participants could repeat the distinction between remembering and knowing. A short distractor task (arithmetic) was used to bring the retention interval out to 5 min for each participant tested. During the test phase, the participants responded to individual words by using one of three keys labeled *remember*, *know*, and *new*. They

were asked to make their judgments as quickly as they could without sacrificing any accuracy.

Results and Discussion

Unless specified otherwise with a p value, the probability of a Type I error did not exceed the conventional 5% in this and the subsequent experiments. Because of the large amount of data generated in these experiments, we used the following conventions for this experiment and those that follow. We analyzed first overall hits and false alarms, without regard to the R and K claims, in order to assess the presence of a mirror effect from context variability. We did this by pooling over low and high word frequency. We did not analyze the word frequency mirror effect, because of its ubiquity under the learning conditions used here. We analyzed next the R and K claims for low versus high context variability and low versus high word frequency items in order to determine where the hit rate and false alarm rate advantages were localized. These analyses were performed separately for the hits and the false alarms. We also combined the hit and false alarm rates into dual-process measures of recollection and familiarity. For recollection, this measure has usually been just the proportion of hits labeled R minus the proportion of false alarms labeled R. For familiarity, we adopted the independence assumption (i.e., IRK) and transformed the K responses, using the $[K/(1-R)]$ approach. These IRK measures are reported as an alternative way to interpret the K responses as reflecting the outcome of a familiarity process. However, the IRK responses were used to derive d' estimates of familiarity accuracy, and in order to facilitate comparisons with recollection accuracy, we calculated d' estimates for recollective accuracy as well. To be clear, the d' estimates are calculated as $[z(H) - z(FA)]$, but for familiarity the IRK measures were used for hits and false alarms, whereas for recollection the proportions of hits and false alarms labeled R were used (see Kishiyama & Yonelinas, 2003).

The data as proportions are summarized in Table 1. The overall hit rate was greater for low context variability items (.80) than for high context variability items (.74), and the false alarms were greater for high (.36) than for low (.27) context variability items [interaction, $F(1,30) = 53.40$]. Thus, context variability yielded a mirror effect, as it did in Steyvers and Malmberg's (2003) experiment. This overall hit rate was analyzed in a 2 (context variability) \times 2 (word frequency) ANOVA model that yielded main effects of context variability [$F(1,30) = 8.55$] and word frequency [$F(1,30) = 9.88$] and a significant interaction [$F(1,30) = 6.26$]. With both word frequency and context variability, the lower class of items was better recognized than the higher class of items, and context variability had its greatest effect when words were high-frequency ones. These effects were driven by a greater degree of recollection, as indexed by raw proportions of R responses, which were more frequent to both low context variability items [$F(1,30) = 59.02$] and low word frequency items [$F(1,30) = 34.25$]. (Throughout, when an interaction is not

Table 1
Overall Recognition, Remember, and Know Proportions and
IRK Estimates of Familiarity as a Function of Context Variability
and Word Frequency in Experiments 1–4

Measure	Word Frequency	Hits			False Alarms		
		Context Variability		Effects	Context Variability		Effects
		Low	High		Low	High	
Experiment 1: 2,000-msec Study							
Overall	Low	.80	.78		.21	.33	
	High	.79	.70	W, C, I	.33	.40	W, C, I
Remember	Low	.58	.43		.09	.10	
	High	.49	.30	W, C	.13	.12	W
Know	Low	.22	.36		.12	.23	
	High	.31	.40	W, C	.20	.28	W, C
IRK know	Low	.54	.63		.13	.26	
	High	.59	.58	none	.24	.32	W, C
Experiment 2: 500-msec Study							
Overall	Low	.78	.72		.32	.42	
	High	.74	.67	W, C	.47	.51	W, C
Remember	Low	.49	.34		.14	.14	
	High	.40	.28	W, C	.19	.19	W
Know	Low	.30	.37		.18	.28	
	High	.33	.39	C	.28	.33	W, C
IRK know	Low	.61	.59		.22	.32	
	High	.56	.57	none	.35	.41	W, C
Experiment 3: 250-msec Study							
Overall	Low	.72	.66		.35	.37	
	High	.71	.70	C	.45	.53	W, C
Remember	Low	.42	.29		.13	.11	
	High	.35	.31	C, I	.15	.15	W
Know	Low	.30	.37		.21	.26	
	High	.36	.39	C	.30	.38	W, C
IRK know	Low	.54	.53		.26	.30	
	High	.56	.57	none	.36	.44	W, C
Experiment 4: Modality Change							
Overall	Low	.75	.72		.24	.34	
	High	.71	.70	W	.32	.42	W, C
Remember	Low	.51	.39		.08	.10	
	High	.42	.32	W, C	.10	.14	W, C
Know	Low	.24	.34		.16	.24	
	High	.29	.37	W, C	.21	.28	W, C
IRK know	Low	.48	.57		.17	.27	
	High	.51	.56	C	.24	.32	W, C

Note—Significant statistical effects are denoted with a W for word frequency, a C for context variability, and an I for interaction. IRK, independence assumption for recollection and familiarity.

reported, this indicates that it was not statistically significant, as in the present case.) Thus, the hit rate advantage due to context variability was localized to better recollection.

In those item classes for which R responses were greater, K responses were issued less frequently, thereby creating a trade-off in R and K claims. There were more K responses to both high context variability [$F(1,30) = 30.24$] and high word frequency items [$F(1,30) = 11.48$]. Converting these K responses to estimates of the contribution of familiarity to recognition memory (“IRK know” in Table 1) yielded no significant differences in the 2×2 ANOVA, although there was a marginal interaction between the two factors [$F(1,30) = 3.78, p = .06$]. If it is real, this interaction suggests that less familiarity is present when recollection is greatest (i.e., the cell with low-

frequency and low context variability items). Therefore, we conclude conservatively that context variability and word frequency were independent factors that contributed solely to increasing recollection and that there was little difference in the contribution of familiarity among the four item classes in this experiment.

We used an identical set of 2×2 analyses to examine the false alarm rates. The overall false alarm rates, without regard to claims of remembering and knowing, yielded a significant effect of context variability [$F(1,30) = 26.35$] and of word frequency [$F(1,30) = 31.45$]. As was expected, false alarms were greater for either high context variability or high word frequency. False alarms claimed to be R responses are generally rare and are often claimed to be true false memories (e.g., Roediger & McDermott,

1995). In the present case, there was no difference in R claims as a function of context variability [$F(1,30) < 1$], but there was a small and significant effect of word frequency [$F(1,30) = 8.80$]. More high-frequency new items received an R response than did low-frequency items. As was expected, the majority of the false alarms were labeled as K responses. More high-frequency distractors were claimed to receive K responses, as compared with low-frequency distractors [$F(1,30) = 16.31$], and more high context variability items received K responses than did low context items [$F(1,30) = 26.34$]. In addition, these two main effects were still present when the IRK correction was used to estimate the contribution of familiarity [smaller of the two $F(1,30)$ s = 21.87].

We conclude from this experiment that the hit rate advantage from both word frequency and context variability is localized in recollection and that the high-frequency and context variability mirrors in false alarms is due to familiarity. The word frequency effects largely replicate the results of Gardiner and Java (1990), but not entirely. They found only changes in recollection due to word frequency and no change in K responses. We found a change in K response proportions for both hits and false alarms, but the contribution of familiarity after correction was

significant only for false alarms. Because Gardiner and Java published their article long before the IRK correction was introduced, they did not analyze their data in the same fashion as we did here. The notable point is that the context variability effects are novel insofar as they suggest that context variability can behave like word frequency, in an additive fashion for recollection of old items and for familiarity with new items.

To verify that these patterns are present in the strict dual-process estimates, we calculated the estimates of recollection and familiarity described earlier, which are summarized in Table 2. As in the uncorrected portion of R responses, only main effects of word frequency [$F(1,30) = 38.55$] and context variability [$F(1,30) = 45.88$] were found for recollection accuracy computed as proportions. The same was true for a translation of recollection to d' values [smaller of the two main effects, $F(1,30) = 32.59$]. For estimates of familiarity, there was a main effect of context variability [$F(1,30) = 7.57$] and one for word frequency [$F(1,30) = 4.19$]. These d' measures in Table 2 reflect the accuracy of the recollective and familiarity components of recognition memory. The best discrimination occurs for low context variability and low word frequency, whereas the worst discrimination occurs when items are high on both factors. This outcome is driven by the hits called R responses in the recollection d' measure and by the false alarms called K responses in the familiarity d' measure of accuracy.

We conclude that context variability acts like word frequency, insofar as the hit rate portion of the mirror effect is based on a greater contribution of recollection for items having low normative context variability, rather than for items that have been preexposed in many different contexts. By contrast, having many preexisting associated contexts causes more false alarms than does having fewer, and that effect is due to familiarity. Moreover, the reader should note that because context variability and word frequency were controlled and orthogonally crossed, the word frequency and context variability effects were independent of one another, as well as being generally additive to one another in a qualitative sense. In sum, context variability behaves like word frequency in the recollective and familiarity processes underpinning recognition memory. We will hold in abeyance any discussion of why recollection should covary with context variability until after the empirical results from all four experiments have been reported.

EXPERIMENTS 2 AND 3

The declaration that context variability acts like word frequency cannot be asserted from Experiment 1 alone. Context variability must also respond to an independent manipulation in the same way as the word frequency dichotomous variable. Having a very short study time has been shown to eliminate the word frequency effect or, at least, change the mirror pattern that is usually found. For example, Malmberg and Nelson (2003) eliminated the word frequency effect in hit rates with a 0.25-sec study

Table 2
Estimates of Recollection and d' Accuracy for Recollection- and Familiarity-Based Responding

Measure	Word Frequency	Context Variability		Effects
		Low	High	
Experiment 1: 2,000-msec Study				
Recollection ($\text{Hit}_R - \text{FA}_R$)	Low	.50	.33	W, C
	High	.36	.18	
As d'	Low	1.79	1.26	W, C
	High	1.24	0.76	
Familiarity d'	Low	1.21	0.94	W, C
	High	1.08	0.75	
Experiment 2: 500-msec Study				
Recollection ($\text{Hit}_R - \text{FA}_R$)	Low	.35	.20	W, C
	High	.21	.09	
As d'	Low	1.23	0.74	W, C
	High	0.74	0.36	
Familiarity d'	Low	1.24	0.79	W, C
	High	0.63	0.43	
Experiment 3: 250-msec Study				
Recollection ($\text{Hit}_R - \text{FA}_R$)	Low	.29	.18	W, C
	High	.20	.16	
As d'	Low	1.01	0.76	W, C
	High	0.71	0.63	
Familiarity d'	Low	0.83	0.63	W, C
	High	0.58	0.32	
Experiment 4: Modality Change				
Recollection ($\text{Hit}_R - \text{FA}_R$)	Low	.42	.28	W, C
	High	.31	.18	
As d'	Low	1.55	1.11	W, C
	High	1.14	0.73	
Familiarity d'	Low	0.93	0.97	none
	High	0.80	0.69	

Note—Significant statistical effects are denoted with a W for word frequency and a C for context variability.

time, and Joordens and Hockley (2000) eliminated it with a lexical decision study sequence in which average reaction times were on the order of 0.6 sec. The argument put forth by Malmberg and Nelson was that short study rates do not allow low-frequency words to attract more attention, whereas Joordens and Hockley made a more general assertion that short study episodes reduce recollection in the hit rate portion of the mirror. If context variability acts like word frequency, very rapid presentation rates should eliminate both the word frequency and the context variability effects that were observed in Experiment 1 with a 2.0-sec study rate. If context variability is somehow more resilient or less resilient than word frequency, an increased study rate might or might not affect recognition performance at two much shorter study rates than those used in Experiment 1.

To test these hypotheses, we conducted two experiments that were identical in their essential properties to Experiment 1. In Experiment 2, we shortened the study rate to 500 msec from the 2,000 msec used in Experiment 1. In Experiment 3, we used a 250-msec study rate, after Malmberg and Nelson (2003). These times were chosen because we did not know a priori whether context variability would be more resilient or less resilient to reduced study time, if it even behaved at all differently from word frequency. We predicted that the 500-msec study rate would preserve the word frequency mirror effect, but if context variability were less resilient, its effects might disappear with this presentation rate, as compared with the results in Experiment 1. By contrast, we expected the word frequency effect in hit rates to be absent at a 250-msec rate, but if context variability is more resilient to shortened study time, the context variability effect could remain at even this very fast study rate. Experiments 2 and 3 will be presented together for the sake of brevity, but we did not perform any statistical comparisons across them.

Method

Participants. Undergraduates from the University of Georgia volunteered in exchange for partial credit toward a course requirement. The participants were tested individually in sessions that lasted approximately 25 min. Twenty-six participants were tested in Experiment 2, and 31 were tested in Experiment 3.

Materials and Procedure. The procedural details of the two experiments were identical to those in Experiment 1, with the exception of how long the words were present in the center of the computer monitor during encoding. As was stated earlier, in Experiment 2 a 500-msec rate was used, and in Experiment 3 a 250-msec rate. With presentation times this short, we inserted a short 100-msec blank screen after the offset of one word and the onset of the next (see Malmberg & Nelson, 2003). Recall that in Experiment 1, subsequent words replaced the word on the previous study trial. In all other respects of collecting R and K responses, the two experiments were identical to the preceding one.

Results and Discussion

We can make the general observation, from Table 1, that reducing study time had the expected effect of attenuating the hit rates and exacerbating the false alarm rates, as compared with the results of Experiment 1 and in comparing the results of Experiments 2 and 3 with one

another. In this respect, the manipulations had their intended effects. To foreshadow, context variability appears to be more resilient to reduced encoding than is word frequency. For brevity, we shall refer to Experiments 2 and 3 as *longer* and *shorter study time*, respectively; and in addition, when two main effects are statistically significant in the 2×2 ANOVAs, we will report only the smaller of the two F values. The context variability mirror effect was present in the longer study time condition, because the hit rate was higher for low context variability items (.76) than for high ones (.70) and more false alarms were made to high context variability items (.47) than to low ones (.39) [interaction $F(1,25) = 39.37$]. For the shorter study time experiment, the hit rate advantage was present for context variability (.72 vs. .68), and the false alarm rate was greater to high context (.45) than to low context (.40) items [interaction $F(1,30) = 17.53$]. To further verify these claims, the longer study time produced main effects of word frequency and context variability in the overall hit rates, without regard to R and K claims [smaller $F(1,25) = 7.48$]. The shorter study time yielded no differences in overall recognition as a function of word frequency, but context variability remained a significant influence [$F(1,30) = 5.14$]. The null result of word frequency for the hit rates replicates the result in Joordens and Hockley (2000), as well as that in Malmberg and Nelson (2003). The significant effect of context variability suggests that this variable is more resilient to reductions in study time than is word frequency.

When the R and K claims for hits were examined, the longer study time yielded main effects of both context variability and word frequency on claims of recollection, as it did in Experiment 1 [smaller $F(1,25) = 9.36$]. With the shorter study time, the word frequency effect disappeared, but the main effect of context variability persisted [$F(1,30) = 20.90$], and there was a significant interaction between the two variables [$F(1,30) = 6.41$]. In brief, the interaction reflects that low context variability still gives rise to more recollection, but the effect is larger with low-frequency than with high-frequency words. Alternatively, the word frequency effect is present only for low context variability items. Given that Joordens and Hockley (2000) found that recollection was reduced for low-frequency items under short study conditions, we largely replicated that effect but found that low normative context variability can still preserve storage of recollective details primarily if the items are also low in word frequency.

The trade-off between R and K responses observed in Experiment 1 is not so neatly ordered in the longer study time condition (Experiment 2). High context variability yielded more K responses than did low context variability [$F(1,25) = 12.72$], but there were no effects of word frequency. By contrast, under the shortest study time conditions, a trade-off was observed between R and K responses as a function of context variability only [$F(1,30) = 9.41$]. More K responses were issued to high context variability items than to low ones. As was found in Experiment 1, the IRK correction to K responses yielded no effect of familiarity on the hit rates, and this was true for both shorter

and longer study durations (all F values were nonsignificant). In brief, both of the two effects were present in recollection at a 500-msec study duration, but the overall word frequency effect was eliminated at a 250-msec rate. At that faster rate, more recollection could still be found, owing to low context variability, but it was found mainly in low word frequency items.

The overall false alarm rate was higher for both high word frequency and high context variability items under longer study time conditions [smaller $F(1,25) = 14.50$]. The same was true with shorter study times [smaller $F(1,30) = 7.29$]. We were somewhat surprised at the amount of recollection being claimed for these false alarms. In the longer study time experiment, there was more erroneous recollection for high-frequency words than for low-frequency words [$F(1,25) = 12.93$]. In the short study time experiment, the significant effect of word frequency became marginally significant [$F(1,30) = 3.74, p = .06$]. However, the magnitude of the numerical difference in R claims is quite small. By contrast, there are robust differences in K proportions for these false alarms, with more being issued in the case of high word frequency and high context variability [smaller $F(1,25) = 13.62$ in the longer study time experiment; smaller $F(1,30) = 12.11$ in the shorter study experiment]. High context variability and high word frequency increased false alarms even when the IRK correction was applied [smaller $F(1,25) = 19.38$ in the longer study time experiment; smaller $F(1,30) = 12.04$ in the shorter study time experiment].

The hits and false alarms were combined to obtain the dual-process estimates of recollection and familiarity (see Table 2). For recollection, low context and low word frequency led to more recollection (based on proportions), as compared with the high classes of those items [smaller $F(1,25) = 23.35$ for the longer study time experiment]. The same pattern of effects was obtained in the translation to d' accuracy values [$F(1,25) = 17.95$]. For the shorter study time experiment, recollection (based on proportions) resulted in the same two main effects [smaller $F(1,30) = 5.71$]. The d' recollection accuracy reconfirmed these two main effects [$F(1,30) = 5.49$]. The return of the word frequency effect on recollection in these dual-process estimates occurred because of the greater erroneous recollection of high-frequency distractors. The d' estimates of familiarity accuracy in the longer study time experiment yielded a main effect of context variability and word frequency [smaller $F(1,25) = 11.82$]. In the shorter study time experiment, the same pattern of results were obtained [smaller $F(1,30) = 6.81$]. Note that both of these d' estimates are greater for low context and low word frequency conditions because recollection in the hit rates increases the recollection-based d' and the greater false alarms to high word frequency and high context variability reduce the familiarity-based d' accuracy.

In summary, Experiments 2 and 3 together were conducted to assess the relative strength of the word frequency and context variability influences on recognition memory. Had the context variability effect been less re-

silient to a study duration manipulation, the outcome of Experiment 2 would have been most different from that of Experiment 1. As it turned out, context variability appears to be a bit more resilient to reductions in study duration, so the outcome of Experiment 3 is most different from that of Experiment 1. Low context variability can still lead to better recollection with only 250 msec of study time, but particularly if the items themselves are also less commonly encountered. If the items are common, low context variability is less efficient at increasing recollective details at such a fast presentation rate. One reason for the difference between low- and high-frequency words could be that context variability has a larger range at low word frequency than it does at high word frequency. We will turn now to a final examination of the influence of these two factors on recollection and familiarity.

EXPERIMENT 4

In recognition memory tests, a modality change from study to test generally does not reduce performance that much (see, e.g., Roediger, Weldon, & Challis, 1989). More important, it generally does not eliminate the word frequency effect (e.g., Lee, Tzeng, Garro, & Hung, 1978). In a remember-know paradigm, a modality change also does not change the subjective state of awareness very much, unless the study conditions orient the participants toward the perceptual qualities of the words, and then a modality change greatly reduces K responses (Gregg & Gardiner, 1994). For all of these reasons, changing the test modality should have no effect in the present paradigm. However, recall that we found movements in K responses in Experiment 1 that Gardiner and Java (1990) did not find in their manipulation of word frequency in a remember-know paradigm. That fact alone, combined with our intuition that the changing of modalities may slightly reduce access to recollective details or change the assessment of the familiarity of test items, prompted us to consider seriously whether a modality change at test would dissociate the word frequency factor from the context variability factor in any manner.

There are several reasons to believe that it might. Transfer-appropriate processing effects or, more generally, context effects do occur in recognition memory under certain conditions (for a review, see Smith & Vela, 2001). Changing the modality between study and test could affect differently the four stimulus classes being investigated here. Recently, Macken (2002) reported that contextual reinstatement effects in recognition are found only in R responses. On the basis of that work, a modality change from study to test should reduce recollection when it is present and, therefore, should attenuate or eliminate either the context variability or the word frequency effects that are based on recollection. In addition, reducing access to recollective details with a test manipulation may affect familiarity assessment, because partial recollective details often give rise to feelings of familiarity (e.g., Hicks, Marsh, & Ritschel, 2002). Finally, Humphreys and Maguire

(2005) found that recognition for old items presented in a novel context can reduce recollection. In sum, there are good reasons to assess the behavior of context variability with a test manipulation such as changing the modality, and that is what we did in this last experiment.

Method

Participants. Thirty-seven University of Georgia undergraduates volunteered in exchange for partial credit toward a course research requirement. The participants were tested individually in sessions that lasted approximately 25 min.

Materials and Procedure. The procedural details of this experiment were identical in all respects to those used in Experiment 1 (2-sec encoding per item), with the exception of the test phase. The only change was to present a given test item auditorily over headphones to the participant, who then proceeded to make a remember-know-new judgment in a fashion otherwise identical to that in Experiment 1.

Results and Discussion

We will follow the convention from Experiments 2 and 3 of reporting only the smaller F value when two main effects were statistically significant. The modality change eliminated the hit rate advantage enjoyed by low context variability items (.73), as compared with high context variability items (.71), and thus, the hit rate portion of the context variability mirror effect was missing [$t(36) = 0.89$, n.s.]. The false alarm rate was greater for high context variability items (.38) than for low context items (.28). Therefore, the interaction was significant, but only because of the absence of a hit rate difference between low and high context variability items [$F(1,36) = 25.22$]. In the overall hit rate, low word frequency increased recognition memory [$F(1,36) = 5.57$], but there was no effect of context variability. Nevertheless, more R responses were reported for low-context and low-frequency items than for the high classes of items [smaller $F(1,36) = 13.64$]. There was a trade-off in R and K responses, because K responses increased as a function of high versus low context variability and as a function of high versus low word frequency [smaller $F(1,36) = 6.13$].

In a departure from the previous three experiments, correcting the K responses showed that the estimate of familiarity was not invariant among item classes but, rather, was greater for high context variability items than for low items [$F(1,36) = 7.27$]. Consequently, the modality change allowed familiarity differences to be expressed as a function of the difference in context variability. Perhaps the marginal interaction found in Experiment 1 for these estimates was corroborated here, insofar as low context variability increases recollection and reduces recognition based on familiarity. The reader should note that overall performance in this experiment is comparable not to performance in Experiment 1, but rather to that in Experiment 3, which had 250-msec encoding per item. Therefore, performance was lower due to the modality change, but the pattern of recollection (Rs) for old items claimed to be old preserved both the word frequency and the context variability effects observed in Experiment 1. Because the overall mirror effect was absent in the hits but recollection was sensitive to both context variability and word

frequency, the modality change at test greatly attenuated recollection but did not eliminate the effects of frequency and context variability on recollection. In addition, some of the losses in recollection due to the modality change were compensated for by increases in familiarity, at least as a function of context variability.

The overall false alarm rate without regard to claims of R versus K responses was greater for both high context variability and high word frequency [smaller $F(1,36) = 14.95$]. Higher rates of false recollection (Rs) were once again obtained than would be anticipated, with significantly more of them occurring in high-frequency words and high context variability items [smaller $F(1,36) = 6.59$]. As was stated before, these are very small numerical effects that probably do not warrant much attention. As was anticipated, both main effects were found in K claims [smaller $F(1,36) = 6.58$]. These patterns are the same in the IRK corrected measure [smaller $F(1,36) = 9.54$].

When the hits and false alarms were combined into dual-process estimates, both low context variability and low word frequency significantly increased recollection [as proportions; smaller $F(1,36) = 21.36$]. This outcome was true in the translation to d' accuracy values as well [smaller $F(1,36) = 21.28$]. The d' values for familiarity accuracy are statistically indistinguishable from one another (all F s nonsignificant). What appears to be happening when the modality is changed from study to test is that recollection is attenuated in the hit rates to studied items but low context variability and low word frequency are still independently exerting an influence on retrieving recollective details. The modality change has also affected the assessment of familiarity for both old and new items. The important point is that changing the modality from study to test has affected access to recollective details and also has changed how people assess familiarity relative to a list of studied items. We now will consider more generally what implications these experiments have for recognition memory.

GENERAL DISCUSSION

Normative context variability reflects the average number of preexperimental contexts in which various concepts occur. A concept such as *dinosaur* arises in very few contexts, whereas a concept such as *tragedy* is associated with many different everyday contexts. Both of these concepts are of low word frequency, but they and items like them gave rise to different amounts of recollection when previously studied and different amounts of familiarity when they went unstudied. Declaring that context variability acts much like word frequency, as we have done here, does not specify whether the greater recollection that accrues to low context variability items has exactly the same locus as studied low word frequency items. Theories accounting for the low word frequency advantage include increased attention, increased variability in the encoding experience, and greater recollection associated with low-frequency items. We could assert that low context variability acts by any one, several, or all of these same mechanisms that

confer the word frequency effect. And it may. However, in what follows, we will take a slightly different tack and will offer a potentially novel explanation for the differences in recollection and familiarity that was observed in the four experiments presented herein.

Items that occur in very few preexperimental contexts may have fewer associations in memory, as compared with high context variability items. In the examples just given, the concept of *tragedy* may be associated with a divorce, a train accident, a case of cancer, a flood, or a suicide. In this way, the concept is likely to have multiple context associations, as well as other kinds of associations. By contrast, the concept of *dinosaur* may be associated with a movie or a museum. Couched in terms of spreading activation, high context variability items have a larger fan than do low context variability items. When they are studied, items with high context variability may access a greater number of associated memories, each of which becomes activated only slightly. When a low context variability item is studied, its associative fan is small, and so only a very few associated memories are activated, but the activation accruing to each of them is greater. Following this logic, the greater recollection associated with low context variability items may occur because people are activating more idiosyncratic associations to low context variability items.

There is converging evidence to suggest that this, indeed, could be the locus of the effect. Gardiner, Ramponi, and Richardson-Klavehn (1998) analyzed a corpus of transcripts detailing the reasons that the participants gave for issuing R versus K responses. They found that recognition associated with recollection was driven, in part, either by reports of specific elaborative mnemonic techniques during encoding or by involuntary reminders in relation to the self. The advantage conferred by low context variability could be due to these idiosyncratic reminders that occur during the study sequence (e.g., remembering a television station replaying the movie *Jurassic Park* upon seeing the word *dinosaur*). Because high context variability items have a larger fan, we are suggesting that these involuntary specific reminders occur less frequently and so give rise to threshold recollection somewhat less often (see also Nelson, Zhang, & McKinney, 2001).

According to this account, there is a direct association between the concept and the specific memory that gives rise to recollection. Of course, this need not be so. The association could be mediated by the context(s) associated to the item. For example, *dinosaurs* may make one think of grade school, which in turn may trigger a recollection of a field trip to a dinosaur park. The end result, however, is the same—namely, a concept with fewer associations has a higher probability of triggering more specific memories, which leads to recollection during the recognition test. The word frequency effect could be driven by the same or similar mechanisms (but see Nelson & McEvoy, 2000). Words that are low in their normative frequency may have fewer but more specific associations than do words that are encountered more frequently, as is implied by theoretical models such as SAM (Gillund & Shiffrin, 1984).

The concept of *length* is high in word frequency and may not have any particular associations, whereas the concept of *burial*, which is low in word frequency, may make one think of the last funeral one attended.

This account bears some resemblance to Reder et al.'s (2000) source of activation confusion (SAC) theory, which was offered to account for word frequency mirror effects. The basis of the SAC model is that the size of the associative fan influences recognition memory judgments. For example, Reder, Donavos, and Erickson (2002) manipulated the number of different fonts that words were presented in. Words that shared fonts with other words were recognized less frequently. Thus, to the extent that font style is a context attribute, sharing additional associations to other words hurt performance. Because we found that context variability and word frequency did not always behave in the same manner, that may pose a problem for the SAC model. However, the main difference between the account offered here and the SAC account is that we have assumed that some idiosyncratic recollection may be part of the underpinning of the context variability effect, whereas SAC is relatively silent on the nature of the fan.

Alternatively, low word frequency may act like novelty or, perhaps, even may produce something akin to the isolation effect (e.g., Hunt & Lamb, 2001). Rare words may be processed in a manner in which they are noticed to be different from other items in their surrounding context or to have a relatively distinctive orthography. Our point is that although context variability behaves much like word frequency, the basis for greater recollection in each of the low classes need not be the same. Our experiments do not allow us to address this question, but until studies can be conducted that specifically address this issue, a defensible stance is that the bases of recollection (and/or familiarity) could be different for the two effects.

The familiarity measured in false alarms must be owing to preexperimental differences in prior processing. For high context variability items, the activation, during test, of many weak context associations may be misinterpreted as familiarity accruing from the item's having been studied. Because these items evoke less recollection when they are studied, and because they evoke fewer specific idiosyncratic reminders, these also do not occur during testing. Consequently, participants less often have a specific recollection associated with high context variability items that would allow them to use something akin to the recall-to-reject strategy for these high context variability items (cf. Rotello, Macmillan, & Van Tassel, 2000). By contrast, a recall-to-reject mechanism could occur more frequently with low context variability distractors if specific associations come to mind, during the test, that a participant knew would have been remembered if the item had been studied. This account is similar in spirit to the argument that participants can reject distractors on the basis of metacognitive information about what should be contained in memory after studying a list of words (e.g., Brown, Lewis, & Monk, 1977; Strack & Bless, 1994). Alternatively, the fewer associations in these items give rise to smaller amounts of assessed familiarity. Of course,

the differences in false alarm rates could be a function of some combination of the ability to recall to reject and a familiarity assessment. To the extent that IRK correction estimates the contribution of familiarity, a recall-to-reject explanation cannot be the only locus of differences in the false alarm rates for low versus high context variability items. We only raise the recall-to-reject issue because the patterns in the claims of (false) recollection for false alarms, although numerically small, are generally consistent with the present analysis.

The outcomes from reductions in study time and the change in modality from study to test are also consistent with the foregoing analysis. Reductions in study time are likely to reduce the opportunity for specific, idiosyncratic associations to come to mind. In Experiment 3, we found that a very short study time eliminated the overall word frequency advantage in recognition memory. Context variability was still a significant factor contributing to recollection, but mainly when concepts were also of low word frequency. The rarity of a concept may contribute to the probability that studying a low context variability item will evoke a specific, idiosyncratic recollection. Or the rarity of a concept may contribute to the speed with which an association is recollected. Our experiments do not speak to these conjectures, but they may provide directions for future work. Nevertheless, the change in modality from study to test was performed because of our intuition that it could reduce access to the recollective details established during encoding. The hit rate advantage due to context variability was indeed eliminated, the contribution of recollection due to context variability was preserved but attenuated, and a greater contribution of familiarity in high context variability items emerged. The appearance of the familiarity factor, which was foreshadowed in the marginal interaction in Experiment 1, may occur in hit rates for high context variability items for the same reason it always occurs in false alarms. With a modality change, the participants realize that they no longer have a perfect copy cue of what they studied originally, and consequently, they increase their reliance on familiarity.

One potential criticism of the foregoing analysis concerns low context variability items that are also of high word frequency. We have asserted that the associative fan is smaller for items that appear in few contexts, so the argument could be made that repeatedly strengthening those few context associations should likewise increase the probability that any idiosyncratic reminders will occur, thereby increasing recollection for items of low context variability but of high word frequency. One rebuttal to this argument is that concepts that are encountered frequently undergo what Tulving (1972) called *decontextualization* (see also Conway, Gardiner, Perfect, Anderson, & Cohen, 1997). That is, they exhibit a remember to know shift insofar as they lose the ability to be associated with any particular context. In terms of the theory being advocated here, high-frequency words have many associations beyond just context associations, and the greater number of these other associations may act analogously to having many context associations. Therefore, context associations do

matter, but we are not denying that other associations also influence the bases of recognition memory.

Before closing, we will briefly note that some readers may have found our contrasting of word frequency against context variability to be a red herring approach. Rather, the role of context variability may speak to fundamental theories of memory. Currently, Dennis and Humphreys's (2001) BCDMEM model has critics who have their own models that make somewhat different assumptions (e.g., Howard & Kahana, 2002). Perhaps, in combination with criticisms of the remember-know procedure, we could have recast those R and K judgments as simple differences in confidence levels. If this is done, the present results may help to discriminate eventually between current models of memory. Although we have not attempted such an analysis here, the empirical outcomes from this study could be important in that regard.

Context variability has only recently been dissociated empirically from word frequency by Steyvers and Malmberg (2003). Prior to this advancement, word frequency and context variability were shown to be highly correlated naturally (e.g., Dennis & Humphreys, 2001). We have demonstrated that both variables give rise to similar patterns of recollection and familiarity but that context variability may behave in qualitatively different ways in response to independent manipulations. We have investigated only reduced study time and changes in the test modality. Given that we encounter concepts that have different numbers of preexperimental contexts in everyday life, understanding how that factor affects later recognition memory is a potentially important endeavor. One may interact with a colleague quite frequently at work but never see that person outside of one's workplace. By contrast, holding frequency of encounter constant, one could see a different colleague at work, in the grocery store, at a chain store shopping, and so forth. Memory for one particular encounter of the first colleague could be better than a similar memory for the second colleague. Clearly, testing such an idea represents a direct extension of the present study, but the data herein indicate that recognition memory might be different in these two cases. Another direction for future work could be to examine whether context variability affects the decontextualization of items repeatedly experienced in the same versus different experimental contexts. Still additional work could explore whether context memory itself (i.e., source monitoring) is affected by preexperimental context variability. In brief, much remains to be learned about context variability, but we have at least provided some initial evidence that low context variability increases recollection in studied items and increases familiarity in unstudied items.

REFERENCES

- ARNDT, J., & REDER, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 830-842.
- BROWN, J., LEWIS, V. J., & MONK, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology*, *29*, 461-473.

- CARY, M., & REDER, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory & Language*, *49*, 231-248.
- CONWAY, M. A., GARDINER, J. M., PERFECT, T. J., ANDERSON, S. J., & COHEN, G. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393-413.
- DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- GARDINER, J. M., & JAVA, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23-30.
- GARDINER, J. M., RAMPONI, C., & RICHARDSON-KLAVEHN, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness & Cognition*, *7*, 1-26.
- GARDINER, J. M., & RICHARDSON-KLAVEHN, A. (2000). Remembering and knowing. In E. Tulving and F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244). New York: Oxford University Press.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20.
- GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 5-16.
- GREGG, V. H., & GARDINER, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology*, *6*, 131-147.
- GUTTENTAG, R., & CARROLL, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory & Language*, *37*, 502-516.
- HICKS, J. L., & MARSH, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review*, *6*, 117-122.
- HICKS, J. L., MARSH, R. L., & RITSCHER, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 503-508.
- HIRSHMAN, E., & ARNDT, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 1306-1323.
- HOCKLEY, W. F. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, *22*, 713-722.
- HOWARD, M. W., & KAHANA, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269-299.
- HUMPHREYS, M. S., DENNIS, S., MAGUIRE, A. M., REYNOLDS, K., BOLLAND, S. W., & HUGHES, J. D. (2003). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 797-812.
- HUMPHREYS, M. S., & MAGUIRE, A. M. (2005). Recollection and familiarity: Redundancy at the item level. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and applications. The 4th Tsukuba International Conference on Memory* (pp. 61-84). Mahwah, NJ: Erlbaum.
- HUNT, R. R., & LAMB, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 1359-1366.
- JACOBY, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 3-26.
- JACOBY, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 3-22.
- JACOBY, L. L., YONELINAS, A. P., & JENNINGS, J. M. (1997). The relationship between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13-47). Mahwah, NJ: Erlbaum.
- 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, & Cognition*, *26*, 1534-1555.
- KELLEY, C. M., & JACOBY, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, *98*, 127-140.
- KISHIYAMA, M. M., & YONELINAS, A. P. (2003). Novelty effects on recollection and familiarity in recognition memory. *Memory & Cognition*, *31*, 1045-1051.
- LEE, A. T., TZENG, O. J. L., GARRO, L. C., & HUNG, D. L. (1978). Sensory modality and the word-frequency effect. *Memory & Cognition*, *6*, 306-311.
- MACKEN, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 153-161.
- MALMBERG, K. J., & NELSON, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*, 35-43.
- MCCLELLAND, J. L., & CHAPPELL, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724-760.
- NELSON, D. L., & MCEVOY, C. L. (2000). What is this thing called frequency? *Memory & Cognition*, *28*, 509-522.
- NELSON, D. L., ZHANG, N., & MCKINNEY, V. M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 1147-1159.
- REDER, L. M., DONAVOS, D. K., & ERICKSON, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition*, *30*, 312-323.
- REDER, L. M., NHOUYVANISVONG, A., SCHUNN, C. D., AYERS, M. S., ANGSTADT, P., & HIRAKI, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 294-320.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803-814.
- ROEDIGER, H. L., III, WELDON, M. S., & CHALLIS, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.
- ROTELLO, C. M., MACMILLAN, N. A., & VAN TASSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, *43*, 67-88.
- SMITH, S. M., & VELA, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203-220.
- STEVENS, M., & MALMBERG, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 760-766.
- STRACK, F., & BLESS, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory & Language*, *33*, 203-217.
- TULVING, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.
- YONELINAS, A. P., & JACOBY, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language*, *34*, 622-643.

NOTE

1. Throughout this article, we appeal to a dual-process theory of recognition. We would not argue that a single-process theory could be entirely consistent with the results that we will eventually report. These experiments were not designed to shed any light on the single- versus dual-process issue, and accordingly, we understand that there will be readers who will prefer an alternative explanation for our results.