

The effect of context variability on source memory

RICHARD L. MARSH

University of Georgia, Athens, Georgia

GABRIEL I. COOK

Claremont McKenna College, Claremont, California

and

JASON L. HICKS

Louisiana State University, Baton Rouge, Louisiana

High-context-variability (HCV) items are experienced in many more preexperimental semantic contexts than are low-context-variability (LCV) items. LCV confers an advantage to recognition memory (Steyvers & Malmberg, 2003). In the present study, we tested whether or not that advantage could be caused by enhanced source memory. Both context variability and word frequency were manipulated, and both factors generally affected source monitoring. Accuracy was better for LCV items than for HCV items and better for low- than for high-word-frequency items. We also considered whether context variability exerts its influence at encoding or at retrieval. We concluded that better recognition memory for LCV items was due, in part, to better retrieval of contextual details that specify how an item was originally studied.

A concept is said to have high context variability (HCV) if it appears in many different everyday contexts. By contrast, a concept that appears in only a few contexts is said to have low context variability (LCV). For example, *immunity* and *odor* are LCV and HCV words, respectively. The word *immunity* might appear only in medical or legal contexts, whereas the word *odor* might occur in references to cooking, cleaning, hiking, hobbies, hygiene, tragedies reported on the news, and so forth. Context variability is not the same as word frequency, although these two variables are highly correlated with one another (see, e.g., Dennis & Humphreys, 2001). Steyvers and Malmberg (2003) investigated recognition memory for words from four different classes of concepts derived by orthogonally crossing high word frequency (HWF) and low word frequency (LWF) with HCV and LCV. Not surprisingly, recognition was better for LWF words than for HWF words. However, recognition was also better for LCV words than for HCV words (cf. Cook, Marsh, & Hicks, 2006); the same was true of free recall (Hicks, Marsh, & Cook, 2005). Thus, unlike the word frequency effect, which reverses when tested with recognition and free recall (with pure lists), memory is generally better for LCV items than for HCV items. In the present study, we investigated whether or not

differences in context memory (i.e., source memory) exist as a function of preexperimental context variability.

Because the term *context* sometimes invites confusion in the memory literature, we begin by explicitly defining the constructs that we are using in this article. When we use the term *context variability*, we are referring to the number of semantic preexperimental contexts in which a word has been experienced. More specifically, Steyvers and Malmberg (2003) differentiated HCV and LCV by counting the number of separate texts (on topics such as business, health, language, and arts) in which a word appeared in the Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998) corpus. HCV items are experienced in a larger number of semantic topics than are LCV items. Because the HCV and LCV items we used can be equated for word frequency, we assume that they have been experienced in an equivalent number of visual/perceptual contexts. Because there is no reason to believe that written word frequencies do not mirror spoken word frequencies, we also assume that these classes of items have an equivalent number of auditory contexts as well. Thus, in our stimulus set HCV and LCV items do not correspond to different degrees of preexperimental perceptual context variability. By contrast, the source-monitoring framework, as tested here, pertains to memory for the episodic context surrounding the appearance of a word in the experiments conducted—that is, memory for whether a word was seen, heard, generated from an anagram, and so forth. Thus, the two uses of the term *context* are very different. Nonetheless, our goal was to ascertain whether or not number of preexperimental contexts affects the episodic binding of source information into memory.

We thank Todd Lindsay and Justin Waldorf for their 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: rlmash@uga.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

Our inquiry began with the assumption that context variability behaves somewhat similarly to word frequency. Theories of the LWF advantage in recognition memory include theories of heightened attention during learning (see, e.g., Glanzer & Adams, 1990), increased encoding variability (McClelland & Chappell, 1998), and increased recollective details at both study and test (see, e.g., Guttentag & Carroll, 1997; Joordens & Hockley, 2000). In fact, both the remember-know procedure and process dissociation estimates demonstrate that LWF items are associated with more recollection than are HWF items (see, e.g., Gardiner & Java, 1990; Guttentag & Carroll, 1994, 1997). According to those who believe that recall for the contextual details of a learning episode is based on recollection, source-monitoring performance should be better for LCV than for HCV items if context variability has an effect similar to that of word frequency. Guttentag and Carroll (1997) found that LWF words not only were recognized better than HWF words, as evidenced by a larger proportion of "remember" responses, but also yielded better source-monitoring performance for the gender of the speaker or the type of judgment made during learning.

As additional evidence in favor of our assumption, Meiser and Broeder (2002) found that answering one source question correctly at test helped participants recover source memory for a different attribute, but only when memory was associated with recollection (defined as the ability to correctly produce a "remember" response in the remember-know procedure). There is also evidence that accurate source memory is found primarily when participants report that they remember rather than that they just know (Perfect, Mayes, Downes, & Van Eijk, 1996). Therefore, source memory is improved when recollection is involved. In a study of context reinstatement, Macken (2002) discovered that reinstating contextual details of the study episode (e.g., word color, font) increased recognition memory in "remember" responses but not in "know" responses. Consequently, there appears to be some kind of reciprocal relationship in which increased context information improves recollection-based memory, and recognition memory associated with recollection yields better source memory for low-frequency words.

The foregoing analysis notwithstanding, improved source memory for LCV items is not necessarily preordained. An alternative prediction is that in the source-monitoring framework it is quite clear that partial information that gives rise only to familiarity can serve as an accurate basis for a source-monitoring judgment (see, e.g., Johnson, Hashtroudi, & Lindsay, 1993; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick, Klein, Dodson, & Shimamura, 2000). For example, Hicks, Marsh, and Ritschel (2002) demonstrated that accurate source decisions can be made when familiarity (measured in terms of "know" responses) outweighs recollection (measured in terms of "remember" responses). In addition, even if context variability behaves like word frequency, the increased recollection that might be associated with LCV items could be primarily criterial for recognition and noncriterial for source monitoring. Criterial recollection is recollection that improves the

memory task at hand, whereas noncriterial recollection is idiosyncratic and does not improve one's ability to perform a task (Yonelinas & Jacoby, 1996). In the present case, if LCV, like LWF, gives rise to more recollection, the resulting recollection may aid recognition memory but confer no advantage for source memory judgments.

To elaborate, studying the LCV item *immunity* may prompt the thought of having gotten an influenza shot 3 months ago. On a recognition test, that specific recollection will be criterial insofar as it serves as a recollective detail that the item was indeed studied. If words were studied from different sources (e.g., seen vs. heard), memory for the flu shot could be noncriterial for remembering whether the item was read or heard. Stated in the terminology of the source-monitoring framework, LCV may improve the old-new detection component of a source judgment but leave unaffected the source discrimination component of, say, heard versus seen items. For this reason, and because the question of central interest concerns source memory, the primary dependent measures we used in this study are source scores conditionalized on old-new recognition. The important point of the foregoing analysis was that there are alternative predictions regarding whether or not context variability affects source memory. Given the importance of episodic context information in newer models of memory (see, e.g., Dennis & Humphreys, 2001; Howard & Kahana, 1999, 2002), we believed that it was time to address the degree to which preexperimental context variability affects episodic context memory.

The foregoing analysis is not compromised by the fact that different kinds of information are used in a recognition memory task than in a source memory task. In analyses of the types of information recollected in a recognition task, Bodner and Lindsay (2003) showed that spontaneous recollection of the episodic context (i.e., source memory) occurs frequently. Thus, our claim that greater source memory for LCV than for HCV items would give rise to recollection in a recognition task has been substantiated by those who have identified the bases for recollection in recognition-memory tasks (see also Gardiner, Ramponi, & Richardson-Klavehn, 1998). Our earlier point was that such recollection need not account for any differences between HCV and LCV items on a source test, and not that spontaneous recollection of source information does not occur at all in recognition-memory tasks. What may pose difficulties for the present approach is that we have chosen our sources arbitrarily, and perhaps the sources chosen are somehow special or different from other pairs of sources that we did not choose. We acknowledge that the present approach has a strong empirical motivation, but currently there are few theoretical building blocks with the stimulus materials used here that separate the effects of context variability from those of word frequency, given that they are so strongly correlated naturally. Moreover, the theory we offer later is a general one that should extend to many different kinds of combinations of sources; therefore, we tentatively offer the prediction that the results from the following experiments should indeed extend to other combinations of sources. For this reason, we tested source

monitoring with two external sources in Experiment 1 and with a combination of external and internal sources in Experiment 2, for generality.

In two experiments, we manipulated the sources, word frequencies, and context variabilities of studied items. The word frequency manipulation provided a touchstone to Guttentag and Carroll's (1997) finding that source memory for LWF words is better than that for HWF words. If LCV items confer better source memory than HCV items, then part of what is being recollected may be context information, or it may be associated in some direct way with accessing that information. Alternatively, if the better recognition memory for LCV items confers no source memory advantage, then this finding will be valuable for constraining the theoretical reasons behind Steyvers and Malmberg's (2003) finding that recognition memory covaries with context variability. We postpone our discussion of how context variability would improve source memory until the General Discussion section, after the data from the two experiments have been presented.

EXPERIMENT 1

In the first experiment, seen and heard sources were chosen because they were both external sources, and normative context variability represents the degree to which concepts appear in external contexts or environments. We believed that a case of external-external source monitoring would be an ideal place to begin our investigation. If source memory is better for LCV than for HCV items, then part of the advantage that Steyvers and Malmberg (2003) found for recognition memory may be attributed to recollection of contextual details. If source memory does not change as a function of context variability, then the advantage they found can be assumed to lie in some other aspect of recognition memory. For generality, we tested source monitoring with two different types of judgments. In the simultaneous condition, participants made standard source judgments, choosing "seen," "heard," or "new" on each trial. In the sequential condition, they first made an "old"- "new" judgment and then a "seen," "heard," or "unsure" source judgment. The latter condition was tested to ascertain whether or not participants display more conservative decision criteria under the simultaneous method of testing than under the sequential method.

Method

Participants. Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course research requirement. Each participant was tested individually in sessions that lasted approximately 30 min. Sixty-four volunteers agreed to participate, none of whom had previously served in any source-monitoring experiment. Half of the participants served in the simultaneous condition, and the other half in the sequential condition.

Materials and Procedure. The materials were taken from the appendix provided by Steyvers and Malmberg (2003). We pared their list down slightly to arrive at 268 words, with 67 items in each of four stimulus classes generated by crossing LWF and HWF with LCV and HCV. The original list has 73 items per class; we assumed that our subset had the same fundamental characteristics as the original. The details of how these items were generated and the descrip-

tive statistics concerning the items can be found in Steyvers and Malmberg. For each participant, the software randomly chose 15 of the 67 items in each stimulus class to be presented by the heard source; it did the same for items to be presented by the seen source. Thus, the participants studied 60 seen items and 60 heard items. An additional 15 items from each of the four stimulus classes were chosen as "new" items. Therefore, each participant was tested on 180 items, with 60 items from each of the seen, heard, and new sources. To reiterate, each of these three sources included 15 each of LCV-LWF, HCV-LWF, LCV-HWF, and HCV-HWF items. Items were randomly intermingled during both study and test, and this randomization was performed separately for each participant.

Prior to encoding, all the participants were informed that they were going to learn a list of words for an unspecified memory test. They were also instructed that half of the items would be presented in the center of the computer monitor and that they should read them silently to themselves. They were also told that the other half of the list would be spoken by the experimenter. They read these instructions on the computer monitor, and then the experimenter reiterated them. During the study phase for a seen item, a fixation point and a warning tone were presented for 250 msec. Then, the word appeared on the participant's monitor for 3 sec, after which the next study trial began. For the heard items, a 250-msec warning tone was sounded, after which the written word appeared on the experimenter's computer work station, behind the participant. The experimenter read each of these words aloud. No formal distractor task was used; instead, the participants were instructed on how to make source-monitoring judgments, which took approximately 2 min. They were informed that for each test item presented, they were to decide whether it had been seen on the computer monitor, had been heard spoken by the experimenter, or was new. In the simultaneous condition, three labeled keys were provided for this purpose. In the sequential condition, the participants were informed that they had to press one of two labeled keys to indicate whether the item was old or new. They were further instructed that if they chose "old," they would have to indicate the source of the item by pressing one of three labeled keys ("heard," "seen," or "unsure"). The participants were asked to respond as quickly as they could without sacrificing accuracy.

Results and Discussion

Unless specified otherwise by a *p* value, the probability of a Type I error does not exceed 5%. Because there were four item classes embedded within each of the three sources, reporting raw correct and incorrect source attributions as proportions would require the reader to digest an unwieldy 36-48 means in each experiment. Therefore, we have used the data distillation technique of reporting conditionalized source scores and placed the raw scores in the Appendix. For each of the four item classes, the conditionalized score represents the proportion of items labeled "old" that were then identified with the correct source (e.g., heard items correctly labeled "heard" as a proportion of all heard items labeled "old"). Analysis of raw proportions generally yields the same conclusions as analysis of the presented data, and misattributions add little interpretive power to this report (see the Appendix). Before considering the source-monitoring data, we report the inferred recognition hit rates, which account for the proportion of studied items labeled "old" without regard to source-monitoring accuracy.

These recognition data are summarized in Table 1. If the data replicate those of Steyvers and Malmberg (2003) and Cook et al. (2006), then main effects of word frequency and context variability should be present. We analyzed

Table 1
Mean Inferred Recognition Proportions (and Standard Errors) for Experiments 1 and 2

	LCV-LWF		HCV-LWF		LCV-HWF		HCV-HWF	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1								
Simultaneous	.81	.02	.71	.03	.72	.03	.67	.03
Sequential	.78	.02	.69	.03	.75	.03	.64	.03
Experiment 2	.79	.02	.74	.03	.82	.02	.75	.03

Note— LCV, low context variability; HCV, high context variability; LWF, low word frequency; HWF, high word frequency.

the data with a 2 (condition: simultaneous vs. sequential) \times 2 (context variability: low vs. high) \times 2 (word frequency: low vs. high) mixed model ANOVA. As would be predicted from the previous reports, greater recognition was obtained for LWF than for HWF items [$F(1,62) = 22.58$]. In addition, recognition was better for LCV than for HCV items [$F(1,62) = 43.71$]. None of the other five terms in the model reached statistical significance. Thus, simultaneous versus sequential testing does not matter for studied items, and the recognition data replicated the two published reports in which these materials were used.

The conditionalized source-monitoring scores are displayed in Figure 1. As the legend indicates, within each source accurate performance is displayed for the LCV-LWF, HCV-LWF, LCV-HWF, and HCV-HWF items. The data were initially analyzed with a 2 (source: seen vs. heard) \times 2 (word frequency) \times 2 (context variability) \times 2 (condition) mixed ANOVA. Source monitoring was better for LCV than for HCV items [$F(1,62) = 36.38$, $MS_e = 0.02$]. Source monitoring was also better for LWF items than for HWF items [$F(1,62) = 57.35$, $MS_e = 0.02$]. Source and word frequency did interact [$F(1,62) = 6.20$], reflecting the fact that word frequency has a more uniform effect on the heard source than on the seen source. None of the other 12 terms in the model was statistically significant.

Because type of source test (simultaneous vs. sequential) was not significant and did not interact, we pooled these two conditions and analyzed each source separately. Analysis of the seen items yielded main effects of both context variability [$F(1,63) = 19.61$] and word frequency [$F(1,63) = 10.69$], but no interaction [$F(1,63) < 1$, $MS_e = 0.021$]. In other words, context variability affected source memory for the seen items alone. For the heard items, there were also main effects of both word frequency and context variability [$F(1,63) = 38.28$ and $F(1,63) = 18.06$, respectively; $MS_e = 0.027$] and no interaction [$F(1,63) < 1$]. Like the seen items alone, the heard items alone are affected by context variability. Because the omnibus ANOVA revealed the same overall source memory for seen and for heard items, in Figure 1 we provide a third set of bars, depicting the conditionalized source scores (averaged over the two sources), as a more stable representation of overall source-monitoring performance. As can be seen in the figure, the main effects of context variability [$F(1,63) = 31.99$] and word frequency [$F(1,63) = 52.90$] were again not qualified by a significant interaction [$F(1,63) < 1$]. Consequently, no matter how one ex-

amines performance, LCV confers a source-monitoring advantage over HCV.

Source memory for new items cannot be conditionalized as can that for the seen and heard sources. Consequently, the grouped bars in Figure 2 represent the proportions of new items claimed to be new. More new items were claimed to be new for LCV than for HCV items [$F(1,63) = 4.64$, $MS_e = 0.011$]. Source memory was also a function of word frequency, with LWF resulting in better memory than HWF [$F(1,63) = 34.75$]. However, there was an interaction between context variability and method of testing such that the step function within the groups of bars is less uniform in the sequential test condition, reflecting an absence of a context variability advantage for HWF [$F(1,63) = 3.99$]. Finally, more new items were misattributed as having been studied in the simultaneous than in the sequential test condition, which is consistent with the presence of a bias to call items "old" when two of the response options are "old" rather than only one (see Hicks & Marsh, 1999). Therefore, performance for new items has the same overall pattern as the average performance for the seen and heard items. This occurred because the complement to the correct proportions represents false alarms, and false alarms are driven by the different amounts of perceived familiarity of the different classes of items. In this sense, LCV (and LWF) functions to increase the hit rate for studied items and the rejection rate for distractor items. This mirror pattern is consistent with Steyvers and Malmberg's (2003) and Cook et al.'s (2006) recognition memory results.

The only data that have not been addressed pertain to the use of the "unsure" category option in the sequential test condition (see raw proportions in the Appendix). Following the ordering of the grouped bars in Figures 1 and 2, the use of the "unsure" category in conditionalized source scores rises linearly for seen items, from 3.35% in the LCV-LWF case to 15.35% in the HCV-HWF case. For heard items, the analogous rise in conditionalized scores is from 3.41% to 12.38%. Like the analyses of the other dependent variables, analysis of this one led to independent effects of context variability [$F(1,31) = 11.12$] and word frequency [$F(1,31) = 15.02$, $MS_e = 0.010$]. None of the other six terms in the model reached statistical significance.

The overall pattern clearly shows that context variability affects source memory. Items associated with more preexperimental contexts yield worse source monitoring than those with fewer preexperimental context associa-

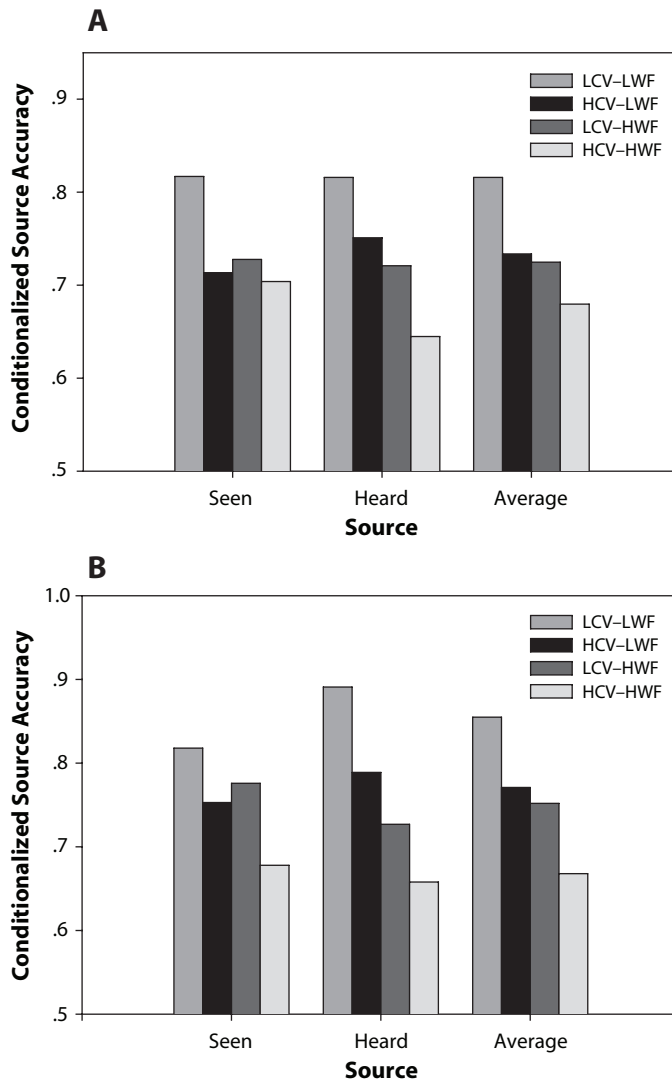


Figure 1. Conditionalized source monitoring scores for old items in Experiment 1. Panel A summarizes the data from the simultaneous condition, and panel B summarizes those from the sequential condition. LCV, low context variability; HCV, high context variability; LWF, low word frequency; HWF, high word frequency.

tions. As we stated earlier, we will provide a theory for this effect in the General Discussion section after we have shown that the effect can be replicated. As Guttentag and Carroll (1997) found, source memory was also better for LWF words than for HWF words. We conclude that some portion of the enhanced recognition that Steyvers and Malmberg (2003) and Cook et al. (2006) found for studied items could be due to recollection of some aspects of the context experienced during learning (cf. Bodner & Lindsay, 2003).

EXPERIMENT 2

We conducted this experiment because it is important to show that context variability has similar effects with

different combinations of sources. Any given source may behave differently when tested with a different source (Marsh & Hicks, 1998). Instead of reusing the seen source as in Experiment 1, we asked participants to generate words from anagrams (Marsh & Hicks, 1998). Anagram generation represents an internal source because it establishes records of cognitive operations involved with solving the anagram. In raw proportions (but not conditionalized scores), source performance is always better for a generated source than for an external source such as seeing the items or hearing them spoken during learning (see the Appendix). Consequently, we predicted that we might disrupt either the word frequency or the context variability effects in the generated source. The important point was to ascertain whether or not context variabil-

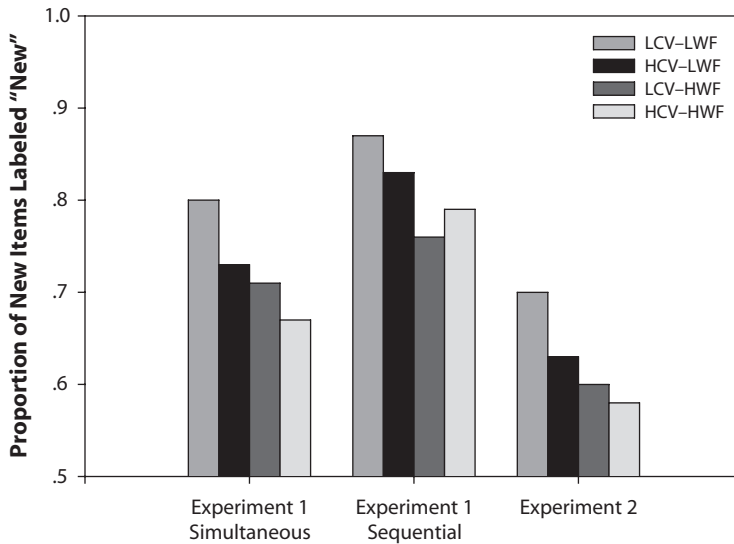


Figure 2. Proportions of new items labeled “new” in both conditions of Experiment 1 and in Experiment 2. LCV, low context variability; HCV, high context variability; LWF, low word frequency; HWF, high word frequency.

ity exerted any influence in an internal–external case of source monitoring.

Method

Participants. Thirty-one undergraduates from the University of Georgia volunteered in exchange for partial credit toward a course requirement. Each participant was tested individually in a session that lasted approximately 30 min. None of the participants had participated in Experiment 1.

Materials and Procedure. In almost every procedural respect, this experiment was identical to Experiment 1. The only difference was that the seen source was replaced with a generated source that we have used on many previous occasions. When an anagram appeared, two letters in the word were interchanged subject to the constraints that the letters not be adjacent, not be identical to one another, and not include the first letter of the anagram. Carets (^) appeared under the two interchanged letters, and the participants were instructed to interchange them mentally, report the word aloud, and press the space bar to initiate the next study trial. They were given as much time as they needed for each word, but in our experience the task takes an average of approximately 3 sec per word (which is why we chose the 3-sec study duration in Experiment 1). At test, the participants pressed one of three labeled keys to indicate whether test items had been generated, had been heard spoken by the experimenter, or were new words. We did not test the sequential condition used in Experiment 1 because of the highly similar results obtained in the simultaneous and sequential conditions.

Results and Discussion

The inferred recognition hit rates are summarized in Table 1.¹ In the 2 (context variability) \times 2 (word frequency) ANOVA, LCV items were recognized as studied more often than HCV items [$F(1,30) = 18.02$], but the word frequency effect was absent [$F(1,30) = 2.66, p > .10$]. The absence of the word frequency effect is probably due to its being trumped by the generation effect. To demonstrate that generation improved memory, when source was entered as a factor in the model it was found

to be statistically significant [$F(1,30) = 97.30, MS_e = 0.032$].

To report the data, we use the conventions summarized in Figure 3. In the 2 (source) \times 2 (context variability) \times 2 (word frequency) ANOVA, main effects were obtained for context variability [$F(1,30) = 16.04$] and word frequency [$F(1,30) = 6.93, MS_e = 0.008$]. There was no overall difference between generated and heard items [$F(1,30) = 1.87, n.s.$]. The absence of a generation effect is common in conditionalized scores because generated items are predominantly misattributed to the weaker (heard) source, whereas heard items are misattributed as new and are not used in the conditionalized analyses. That a generation effect was obtained can be seen clearly in the raw proportions presented in the Appendix. However, these effects were qualified by a significant three-way interaction [$F(1,30) = 7.07$]. As will be shown, this interaction arose because, as in Experiment 1, source memory for heard items was affected by both context variability and word frequency, but generated items were affected only by context variability.

As in Experiment 1, we examined performance for each source separately. LCV generated items were identified more accurately than HCV generated items on the source test [$F(1,30) = 5.94, MS_e = 0.013$]. The word frequency effect was absent [$F(1,30) < 1, n.s.$]. Although the interaction was not significant [$F(1,30) = 2.06$], we tested the simple effects to determine the level of word frequency that yielded the overall significant effect of context variability. Context variability exerted most of its influence on LWF items [$t(30) = 2.31$] and had little influence on HWF items [$t(30) < 1$].

Items spoken by the experimenter at study behaved in much the same manner as those in Experiment 1. LCV improved source memory relative to HCV [$F(1,30) = 10.86, MS_e = 0.008$]. In addition, LWF items led to better source monitoring than did HWF items [$F(1,30) = 18.89$].

The interaction fell in the margin [$F(1,30) = 3.07$]. When performance was averaged together with that for the generated source (third set of bars in Figure 3), source memory for LCV items was better than that for HCV items [$F(1,30) = 17.08$], but there was only a marginal effect of word frequency [$F(1,30) = 3.27, p = .08$] and no interaction [$F(1,30) < 1, MS_e = 0.007$]. The absence of the LWF benefit for generated items is responsible for the overall attenuation of the word frequency influence.

New items correctly labeled “new” (see Figure 2) yielded the same qualitative pattern as they did in Experiment 1; overall performance, however, was worse in this experiment. More new items were attributed to a studied source in this experiment, which commonly happens when the two studied sources differ in relative strength. Participants become willing to call new items “heard” (i.e., coming from an external source) rather than claim that they solved them as anagrams. This general finding has been dubbed the *it-had-to-be-you effect* (Bink, Marsh, & Hicks, 1999; Johnson, Raye, Foley, & Foley, 1981). In any case, source monitoring for new items was better for LWF than for HWF words [$F(1,30) = 14.61$], but the effect of context variability (favoring LCV) was only marginally significant [$F(1,30) = 3.38, p = .07$] and there was no interaction [$F(1,30) = 2.57, MS_e = 0.013$].

As in the first experiment, context variability affected source-monitoring judgments. The presence of fewer preexperimental context associations improved source memory, albeit only for LWF items in the case of generation. For heard items, the effect was present at both levels of word frequency. The fact that normative context variability had any effect at all suggests that memory for how an item was studied could contribute to the differences observed on a standard recognition memory test (Cook et al., 2006; Steyvers & Malmberg, 2003).

GENERAL DISCUSSION

All concepts have many kinds of associations to other concepts in memory, including semantic, phonological, idiosyncratic, spatial, temporal, and of course contextual ones. The present study was concerned primarily with investigating normative context variability, which specifies the degree to which concepts are experienced in fewer versus more preexperimental semantic contexts. Our goal from the outset was to ascertain whether or not enhanced recognition memory for LCV items was due, in part, to better recollection of contextual details associated with a memory trace. Had context variability had only null effects on source monitoring, this study would have shown that the advantage Steyvers and Malmberg (2003) observed in recognition was due to some other contribution to recognition memory, and this would have pruned the search tree considerably. Instead, undeniable influences of context variability resulted from the fact that source-monitoring performance favors LCV items. In fact, it was the somewhat venerated word frequency effect, and not the context variability effect, that disappeared for the generated source in Experiment 2. From this perspective, context variability may have a somewhat stronger influence on memory than word frequency. Given how highly correlated these two variables are naturally, perhaps many of the word frequency effects that have been reported over the years are actually due to context variability or to the joint influence of the two variables acting in concert. Even a cursory visual inspection of Figures 1 and 3 indicates that the first bar in each set of bars, representing the LCV–LWF class of items, consistently represents the best source memory. We assert that LCV can improve memory for how an item was studied and, furthermore, that these contextual details may partially account for the better recognition memory reported by Steyvers and Malmberg.

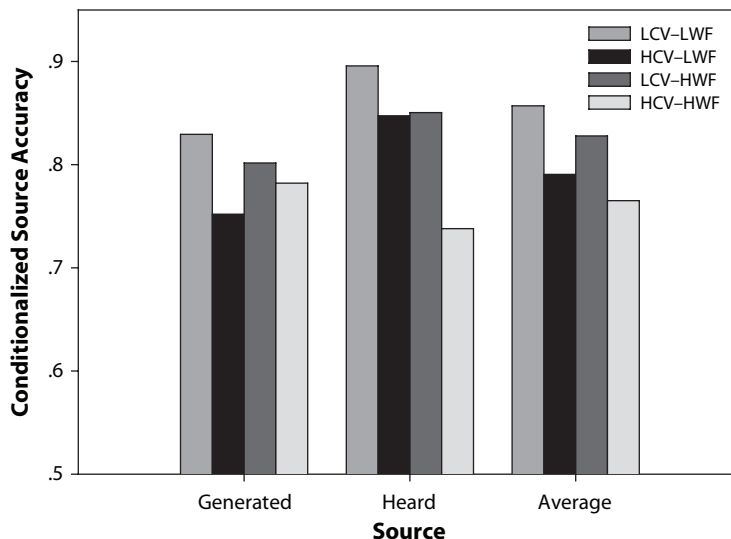


Figure 3. Conditionalized source scores for old items in Experiment 2. LCV, low context variability; HCV, high context variability; LWF, low word frequency; HWF, high word frequency.

The theoretical challenge will be to develop a theory of why context variability affects source memory, or perhaps memory more generally. The processes by which qualitative characteristics (contextual details) are bound into memory are not well specified in the source-monitoring framework (see, e.g., Chalfonte & Johnson, 1996; Johnson & Chalfonte, 1994; Marsh, Hicks, & Cook, 2004). According to theories of attention, attention is required to bind attributes together in both perception and memory (see, e.g., Treisman, 1998). One possibility is that LCV items attract more attention than HCV items, which is one theory for the word frequency effect (see, e.g., Glanzer & Adams, 1990). This explanation may be sufficient to account for increased learning of source-specifying characteristics. Alternatively, this account may need to be augmented with additional assumptions, such as LCV items' having fewer associations in general. Fewer associations may draw to mind less material during encoding and thereby leave more attention available to be focused on the features of the encoding experience. Of course, the locus of the effect need not be the encoding experience at all. Rather, the hypothesized fewer associations with LCV items may make it easier to find and evaluate the source-specifying details during retrieval. By this account, the HCV items have no less source information in them; rather, the greater number of associations may slightly corrupt access to and evaluation of that information.

From the data collected here, the differences in source monitoring as a function of context variability cannot be localized at either encoding or retrieval processes. Therefore, the two theoretical accounts offered in the previous paragraph are to be viewed as tentative at best. To the extent that context variability behaves as word frequency does, all of the theories concerning word frequency favor an encoding explanation. LWF ostensibly conveys an advantage of storage of recollective details at the time of encoding. Some of these details may be idiosyncratic and personal memories, as Gardiner et al. (1998) claimed. This claim may also apply to LCV items, but that cannot be the entire story. Idiosyncratic personal reactions to LCV items would not confer a source-monitoring advantage, nor would similar idiosyncratic reactions to LWF items. Rather, part of the recollective details would also have to include source-specifying details of the encoding experience itself. Until context variability is studied further, all we can conclude is that memory for how an item was studied is affected by the number of preexperimental contexts in which the item occurs. Fewer such contexts leads to better source memory, which we believe can in turn enhance recognition memory. This proposition is consistent with models of recognition memory that emphasize the role of context in enhancing recognition memory (see, e.g., Dennis & Humphreys, 2001; Murnane, Phelps, & Malmberg, 1999).

REFERENCES

BINK, M. L., MARSH, R. L., & HICKS, J. L. (1999). An alternative conceptualization to memory "strength" in reality monitoring. *Journal*

- of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 804-809.
- BODNER, G. E., & LINDSAY, D. S. (2003). Remembering and knowing in context. *Journal of Memory & Language*, **48**, 563-580.
- CHALFONTE, B. L., & JOHNSON, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, **24**, 403-416.
- COOK, G. I., MARSH, R. L., & HICKS, J. L. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition*, **34**, 240-250.
- 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.
- 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.
- GUTTENTAG, R. E., & CARROLL, D. (1994). Identifying the basis for the word frequency effect in recognition memory. *Memory*, **2**, 255-273.
- GUTTENTAG, R. E., & 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., & COOK, G. I. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 1160-1164.
- 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.
- HOWARD, M. W., & KAHANA, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 923-941.
- HOWARD, M. W., & KAHANA, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, **46**, 269-299.
- JOHNSON, M. K., & CHALFONTE, B. L. (1994). Binding complex memories: The role of reactivation and the hippocampus. In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 311-350). Cambridge, MA: MIT Press.
- JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, **94**, 37-64.
- 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.
- LANDAUER, T. K., FOLTZ, P. W., & LAHAM, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, **25**, 259-284.
- MACKEN, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 153-161.
- MARSH, R. L., & HICKS, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1137-1151.
- MARSH, R. L., HICKS, J. L., & COOK, G. I. (2004). Focused attention on one contextual attribute does not reduce source memory for a different attribute. *Memory*, **12**, 183-192.
- 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.
- MEISER, T., & BROEDER, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 116-137.
- MURNANE, K., PHELPS, M. P., & MALMBERG, K. (1999). Context-

- dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, **128**, 403-415.
- PERFECT, T. J., MAYES, A. R., DOWNES, J. J., & VAN EIJK, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology*, **49A**, 797-813.
- QIN, J., RAYE, C. L., JOHNSON, M. K., & MITCHELL, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1110-1115.
- SLOTNICK, S. D., KLEIN, S. A., DODSON, C. S., & SHIMAMURA, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1499-1517.
- STEYVERS, M., & MALMBERG, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 760-766.
- TREISMAN, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London: Series B*, **353**, 1295-1306.
- YONELINAS, A. P., & JACOBY, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness & Cognition*, **5**, 131-141.

NOTE

1. In the review process, we were asked to add these inferred recognition hit rates. Had we chosen to report corrected recognition, the higher false alarm rate to HWF items would have caused the word frequency effect to rematerialize.

APPENDIX
Raw Proportions of Correct Claims and Misattributions
in Experiments 1 and 2

Actual Source of Items and Claims	LCV-LWF	HCV-LWF	LCV-HWF	HCV-HWF
Experiment 1, Simultaneous Condition				
Heard Items				
Heard	.69	.51	.54	.44
Seen	.15	.17	.19	.23
New	.17	.32	.28	.32
Seen Items				
Heard	.14	.22	.20	.20
Seen	.64	.51	.53	.47
New	.21	.27	.28	.33
New Items				
Heard	.12	.16	.15	.17
Seen	.08	.11	.11	.16
New	.80	.73	.73	.67
Experiment 1, Sequential Condition				
Heard Items				
Heard	.73	.58	.57	.45
Seen	.06	.09	.14	.14
Unsure	.03	.04	.07	.07
New	.19	.29	.23	.34
Seen Items				
Heard	.11	.10	.11	.11
Seen	.62	.50	.58	.43
Unsure	.03	.07	.04	.08
New	.25	.33	.28	.38
New Items				
Heard	.08	.08	.13	.08
Seen	.03	.05	.06	.08
Unsure	.02	.04	.04	.05
New	.87	.84	.76	.79
Experiment 2				
Heard Items				
Heard	.62	.51	.63	.46
Generated	.06	.09	.11	.17
New	.32	.40	.26	.37
Generated Items				
Heard	.16	.22	.18	.19
Generated	.75	.65	.71	.68
New	.10	.16	.11	.12
New Items				
Heard	.26	.30	.35	.31
Generated	.04	.07	.06	.11
New	.70	.63	.60	.58

Note—LCV, low context variability; HCV, high context variability; LWF, low word frequency; HWF, high word frequency.