

Context Effects on Remembering and Knowing: The Expectancy Heuristic

David P. McCabe
Colorado State University

David A. Balota
Washington University in St Louis

Three experiments are reported examining the effect of context on remember–know judgments. In Experiments 1 and 2, medium-frequency words were intermixed with high-frequency or low-frequency words at study or at test, respectively. Remember responses were greater for medium-frequency targets when they were studied or tested among high-frequency, as compared with low-frequency, words. The authors proposed a decision-based mechanism called “the expectancy heuristic” to explain why remember responses were more likely when items were studied or tested in the context of words that were relatively less distinct. According to the expectancy heuristic, when items on a recognition test exceed an expected level of memorability they will be given a remember judgment but when they do not, but are still more familiar than new words, they will be given a know judgment. Experiment 3, which varied expectancies about the strength of tested targets, demonstrated the use of the expectancy heuristic, indicating that it operates by selectively influencing the remember criterion rather than by influencing recollection of studied items.

Keywords: context, remember–know, recollection, recognition, signal detection

The remember–know procedure was originally introduced to demonstrate that people could discriminate between the subjective states of awareness associated with auto-noetic and noetic consciousness (Tulving, 1985). Tulving (1985) argued that remember responses were given when subjects could mentally travel back to the moment in time in which they had originally experienced an event, that is, auto-noetic consciousness. In contrast, know responses were given when subjects believed items were studied but they did not experience recovery of contextual information, that is, noetic consciousness.

A plethora of research using the remember–know procedure has revealed systematic effects of different variables on remember and know responses, leading to evidence of functional dissociations of the two responses. The variables that selectively affect remember responses are typically those that encourage more elaborate processing at encoding. For example, deeper levels of processing (Gardiner, 1988), generating studied items (Gardiner, 1988), and more elaborative rehearsal strategies (Gardiner, Gawlick, & Richardson-Klavehn, 1994) all increase remember responses. By contrast, divided attention at study selectively reduces remember judgments (Parkin, Gardiner, & Rosser, 1995), as does the ingestion of sedative–hypnotic drugs, such as midazolam (Huron, Gierisch, & Danion, 2002). Remember responses are also enhanced for certain stimulus manipulations, such as presenting items as pictures as compared with words (Rajaram, 1996) or presenting words of lower frequency (Joordens & Hockley, 2000) or unusual or-

thography (Rajaram, 1998). Other variables, such as masked priming at test, influence know responses but have little effect on remember responses (e.g., Rajaram, 1993). Still other variables, such as increasing study duration, increase both remember and know responses (e.g., Gardiner, Kaminska, Dixon, & Java, 1996).

Remember and know judgments are typically interpreted as indices of different memory systems or processes. For example, dual-process accounts of memory suggest that remember responses are more likely to arise from recollection of past events, whereas know responses, or an estimate based on know responses (i.e., the independence remember–know estimate), are more likely to reflect familiarity-based recognition (Yonelinas, 2002; Wixted & Stretch, 2004). Support for the dual-process account comes from the convergence of different methods of measuring recollection, such that variables that affect remember responses also affect recollection estimates from process dissociation and the shape of receiver operating characteristic curves (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Similarly, Tulving’s (1985) original account of remember–know was based on the notion that remember responses reflected the conscious state of awareness associated with output from the episodic memory system, whereas know responses reflected a state of awareness associated with output from the semantic memory system (see also Gardiner, 1988; Rajaram, 1993; Wheeler, Stuss, & Tulving, 1997).

Context Effects on Remember and Know Judgments

The bulk of experimental work on remember–know judgments has focused on how different encoding and retrieval manipulations affect the reporting of these responses. Less attention has been paid to the decision processes involved in defining what constitutes a remember or know judgment, though obviously decision processes must be important to the use of any subjective report. The present investigation concerns the way in which remember–know judgments are used as a function of different encoding and retrieval contexts, with context referring to the type of items that are

David P. McCabe, Department of Psychology, Colorado State University; David A. Balota, Department of Psychology, Washington University in St. Louis.

We would like to thank Stephanie To and Chelsea Beman for data collection in Experiments 1 and 3, respectively.

Correspondence concerning this article should be addressed to David P. McCabe, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523. E-mail: david.mccabe@colostate.edu

included on the study list or on the recognition test (cf. Bodner & Lindsay, 2003; Davis, Lockhart, & Thomson, 1972). Of course, the effect of encoding and retrieval contexts on memory has been a topic of considerable interest to experimental psychologists for some time (Morris, Bransford, & Franks, 1977; Smith, 1979; Tulving & Pearlstone, 1966; Westerman & Greene, 1998). For example, Tulving and Pearlstone (1966) stated that it

is becoming increasingly clear that remembering does not involve a mere activation of the learned association or arousal of the stored trace by a stimulus. Some sort of a more complex interaction between stored information and certain features of the retrieval environment seems to be involved in converting a potential memory into conscious awareness of the original event. . . (pp. 382–383)

Although encoding–retrieval interactions have been studied extensively since that time, there is little research examining context effects on the subjective experience of remembering, that is, the “conscious awareness of the original event.”

One exception to this paucity of research is a recent investigation showing that the context in which remember–know responses are made can have a strong influence on subjective reports (Bodner & Lindsay, 2003). In Bodner and Lindsay’s (2003) series of experiments, two groups of subjects each studied two lists of words. Each word list was studied using a different orienting task. In the medium with shallow condition, subjects studied one word list using a medium level of processing task (i.e., judging whether objects were commonly used) and studied a different word list using a shallow level of processing task (i.e., determining whether there were any As in the word). In the medium with deep condition, subjects studied one word list using the same medium level of processing task as the first group (i.e., judging whether objects were commonly used), but the other list was studied using a deep level of processing task (i.e., determining whether they would need an item if they were stranded on a desert island). Results indicated that when medium level of processing items were studied and tested with shallow level of processing items, they were more likely to be called old and received more remember responses than when they were studied and tested with deep level of processing items.

As an explanation for their findings, Bodner and Lindsay (2003) proposed a functional account of remembering. The notion is that remember responses are determined by the context in which they are made. According to this account of recollection, retrieval of information supports recollection when it allows subjects to complete the criterial task at hand (Gruppuso, Lindsay, & Kelley, 1997). In the case of Bodner and Lindsay’s study, subjects defined recollection as being based on the type of information that was encouraged by the deeper level of processing task, and, thus, the medium level of processing items received more remember responses when they were tested with the shallow, rather than the deep, level of processing items. The functional account of remembering (Bodner & Lindsay, 2003) highlights the relative nature of remembering, suggesting that remember responses are defined relative to the context in which they are studied and/or tested (see also Norman, 2002).

Another explanation for why some items are remembered and others known, which is compatible with the functional account, is the distinctiveness–fluency framework (Rajaram, 1998; Rajaram & Geraci, 2000). According to this framework, items are given a

remember response when they are perceived as distinctive but are given a know response when they are perceived as fluent but not distinctive. The question, then, becomes why are some items experienced as distinctive and others not? In the present article, we outline an expectancy heuristic, which suggests that information is perceived as more or less distinct and leads to differences in remembering depending on the expected strength of the context in which it is experienced.

The Expectancy Heuristic

We propose an explanation of context effects on remember–know judgments that we call “the expectancy heuristic” to account for Bodner and Lindsay’s (2003) data as well as new data we report in three experiments. The central principle of the expectancy heuristic is that remember responses are given to items that exceed an expected level of memorability, with memorability referring to the likelihood that an item would later be recognized on a recognition test had it been studied. This expected level of memorability is based on the context in which information is studied and/or tested. In Bodner and Lindsay’s study, when the study and test consisted of a mixture of medium and shallow level of processing items, subjects expected the typical memorability of test items to be the average of those two types of items and to be less than the average memorability in the condition in which the medium and deep level of processing items were studied. Because the expected memorability was lower in the medium with shallow condition than in the medium with deep condition, the medium level of processing items were more likely to exceed the expected level of memorability in the medium with shallow condition and therefore were more likely to receive remember responses than in the medium with deep condition.

A quantitative example may help better illustrate this expectancy heuristic. Suppose that, on average, the shallow level of processing items have a memorability of 0.3, the medium level of processing items have a memorability of 0.5, and the deep level of processing items have a memorability of 0.7. In the medium with shallow condition, the expected memorability of test items would be 0.4 (i.e., $[0.3 + 0.5]/2$), and many of the medium level of processing items (average memorability = 0.5) would exceed this expected memorability and would receive remember responses. In the medium with deep level of processing condition, the expected memorability of test items would be 0.6 (i.e., $[0.7 + 0.5]/2$), and few of the medium level of processing items (average memorability = 0.5) would exceed this expected memorability and receive remember responses. Thus, the same items would receive fewer remember responses in the medium with deep condition as compared with the medium with shallow condition because of differences in expected versus actual memorability. It is important to note that the actual memorability of the items need not differ; the expectancy heuristic operates on the basis of the expected memorability, which drives the decision processes involved in remember–know judgments. In the above example, items at different levels of processing differed in memory strength (with strength conceptualized as a combination of recollection and familiarity processes; Wixted & Stretch, 2004), but the actual strength need not differ for context to influence remembering, as we show in the present experiments.

The expectancy heuristic explanation is consistent with the notion that memory judgments are attributional in nature (Geraci & McCabe, 2006; Jacoby & Dallas, 1981; Jacoby, Kelley, & Dywan, 1989; Westerman, Lloyd, & Miller, 2002; Whittlesea & Williams, 2000). For example, according to the discrepancy-attribution hypothesis (Whittlesea & Williams, 2000) context can cause fluency of processing to be discrepant with expected fluency, resulting in changes in old–new recognition judgments. The expectancy heuristic is similar to other attributional mechanisms that have been proposed to explain old–new recognition as well but differs in that we are attempting to explain how people determine what constitutes the experience of recollection or familiarity rather than what constitutes a general experience of “oldness.” Indeed, for subjects to differentiate between remember and know judgments, there must be some experiential difference between items that give rise to a feeling of recollection versus a feeling of familiarity (or knowing). Our expectancy heuristic suggests that the threshold at which subjects make this discrimination is based on the relative memorability within the class of items they believe were studied.

The expectancy heuristic also nicely complements the distinctiveness–fluency framework proposed by Rajaram and colleagues (Rajaram, 1998; Rajaram & Geraci, 2000). According to their framework, items that are perceived as distinct are remembered whereas items that are perceived as familiar, but not distinctive, are known. The expectancy heuristic allows one to operationalize distinctiveness a priori, as a deviation from expected memorability. Thus, studied items that exceed an expected level of memorability will be perceived as distinct and will be given a remember response, and studied items that do not exceed the expected level of memorability will not be perceived as distinct and will be given a know response.

Experiments 1 and 2: Relative Word Frequency

In Experiments 1 and 2, context effects on remember–know judgments were examined using a relative word frequency manipulation (see Figure 1). In Experiment 1, we manipulated relative word frequency during the study phase of a recognition memory experiment. Each of two groups of subjects studied the same series of medium-frequency (MF) words, but for one group those words were studied with high-frequency (HF) context words and in the other group those same MF words were studied with low-frequency (LF) context words. In the example in Figure 1, all subjects studied *lodge* and *house*, but in the HF context, other items on the study list included words like *train* and *house*, whereas in the LF context, other items on the study list included words like *plum* and *cedar*. It is important to note that the tests were identical in both context conditions, consisting of only MF targets and distractors. Thus, the context manipulation was confined to the study phase. Because we know from vast numbers of previous studies that LF words produce better recognition than do HF words, we can be certain that the expected level of memorability would be greater for subjects in the LF study context as compared with subjects in the HF study context. Using the same values as in the previous explanation of Bodner and Lindsay’s (2003) list strength effect, if one assumes that HF words have an average memorability in recognition of 0.3, MF words have an average memorability of 0.5, and LF words to have an average

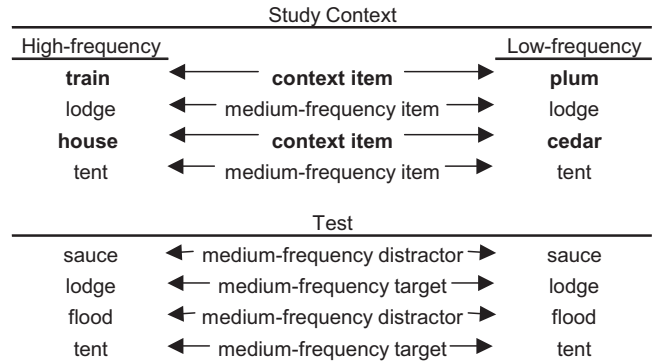


Figure 1. The design of Experiment 1. All subjects studied the same medium-frequency items, but the context items that were studied differed between the low-frequency and high-frequency groups. The study phase illustration represents a run of four possible studied words for each subject, demonstrating that half of the studied items were medium-frequency items and half were context items. The test phase illustration is similar, in that it shows a run of four possible items, half of which were medium-frequency targets and the other half of which were medium-frequency distractors. The test phases were identical for both context groups.

memorability of 0.7, then the expected values of recognition test probes should be 0.4 in the HF context and 0.6 in the LF context. Assuming this average memorability is used as a guide when defining what constitutes a remember response, one would expect MF words to be given more remember responses when their memorability (0.5) exceeds the average memorability of words on the study list (i.e., in the HF context) than when it does not (i.e., in the LF context). It is noteworthy that one cannot simply equate familiarity with memorability, because LF words by definition are relatively less familiar than HF words and yet still produce better recognition memory performance.

We have proposed a decision-based mechanism—the expectancy heuristic—to explain how context affects remember–know judgments, but there exist alternative memory-based explanations as well. A memory-based mechanism would operate to change the strength of representations in memory, as opposed to our decision-based explanation, which suggests that the representations themselves need not change but rather the decision processes involved in evaluating the representations change (see McCabe, Presmanes, Robertson, & Smith, 2004 or Hirshman & Arndt, 1997, for discussions of memory-based and decision-based mechanisms in recognition). Perhaps the most obvious memory-based explanation of how the manipulation of word frequency in the study context might affect remember responses is the attention-likelihood model (Glanzer & Adams, 1990), which suggests that context might influence the way in which items are studied. Studying words of lower relative frequency, for example, studying MF words in a HF context, may bias attention toward the MF words, at least more so than studying MF words in a LF context (e.g., Glanzer & Adams, 1990). If MF items receive more attention during study in the HF context, resulting in more elaborate encoding, then these items would be expected to subsequently receive more remember responses on a later test, along with producing better recognition memory performance. We note, though, that previous studies of the effect of list composition on the word frequency effect have

shown little direct effect of list composition on the hit rate for HF and LF targets (Dorfman & Glanzer, 1988; Malmberg & Murnane, 2002), but this does not rule out the possibility that list composition will affect remember–know judgments. We argue that despite no effect of list composition on overall hit rate, there may still be an effect of list composition on remember responses, which would be consistent with the hypothesis that list composition would influence the perception of distinctiveness at study. That is, LF targets would receive more remember responses in the HF context, where they were perceived as more distinct.

Although results from the study context manipulation could be explained by a version of attention-likelihood theory, this was not true of Experiment 2, in which we examined the effect of relative word frequency (i.e., context) during the recognition test itself. The design was essentially a mirror image of that of Experiment 1 (the study and test phases shown in Figure 1 were reversed). In Experiment 2, all subjects studied a homogeneous list of MF words and were tested on a subset of those MF words. In the HF test context, those MF targets were intermixed with HF (context) distractors, whereas in the LF test context, those targets were intermixed with LF (context) distractors. Although the actual memorability of the studied items did not differ for the test context conditions because the study lists were exactly the same, we predicted that the expected memorability of the test items would differ on the basis of the metamemorial knowledge that LF words are better discriminated than are HF words (Benjamin, 2003; Guttentag & Carroll, 1998). That is, because the LF distractors would be expected to be well remembered had they been studied but the HF distractors would not be expected to be well remembered, the test context manipulation would have the same functional consequence as manipulating the actual memorability of the targets. This idea was well explained by Brown, Lewis, and Monk (1977) in their seminal word frequency effect article: “If the subject judges the item concerned to be of high memorability [i.e., a low-frequency word], the absence of positive memory constitutes stronger evidence against that item than if the subject judges it to be of low memorability [i.e., a high-frequency word]” (p. 470; see also, Strack & Bless, 1994, for a similar idea). Thus, although the MF words were encoded identically in both the HF and LF test contexts, the expected memorability should be greater in the LF test context as compared with the HF test context. As a result, consistent with the predictions for the study context manipulation of Experiment 1, we expected remember responses to be greater in the HF than in the LF test context.

In contrast to the study context manipulation, if test context affected remember–know judgments, then attention-likelihood theory would not provide a compelling explanation of this result. Specifically, if the use of remember–know responses differs as a function of retrieval context, then the results could not be due to differential attentional focus during encoding. Thus, the design of Experiment 2 affords a strong test of the attention-likelihood explanation of the effects of relative word frequency on remember–know judgments.

However, attention-likelihood theory is not the only possible memory-based framework that can explain context effects. It is possible that a contextual manipulation can affect the recollection process itself, regardless of whether context is manipulated at study or test. For example, finding more remember responses for MF targets in the context of HF distractors, as compared with LF

distractors, could be due to higher frequency targets providing better retrieval cues for MF items, perhaps because they co-occur together in more preexperimental contexts than do MF and LF words. A similar mechanism could operate at encoding as well, with higher frequency words providing a richer encoding context than lower frequency words. Thus, it is possible for this recollection change explanation, a memory-based explanation, to account for the study and context effects we are examining using relative word frequency.

Experiment 3: Manipulating Expected Memorability

To adjudicate between memory-based and decision-based explanations for context effects on remember–know judgments, we conducted a third experiment in which the expected memorability of studied items was directly manipulated using slightly different instructions at test. In Experiment 3, subjects studied 60 words, half of which were studied once and the other half of which were studied five times. It is important to note that subjects were only tested on the weak items, that is, the items that were studied once, mixed with distractors. One group of subjects, the expecting strong group, was not told that only the once-presented items would be tested, so they were expecting to be tested for both strong and weak items. The other group, the expecting weak group, was told that only the once-presented items would be tested, so they were expecting to be tested for only weak items. According to the expectancy heuristic, the criterion for remember responses should change on the basis of the expected level of memorability, such that subjects who were expecting weak items would use a more liberal remember criterion and give remember responses to items that include fewer recollective details, as compared with subjects who are expecting strong items. As such, the expectancy heuristic predicts that in the expecting weak group, there should be more remember responses and fewer know responses for both studied and new items, as compared with the expecting strong group. Moreover, as we show in the introduction to Experiment 3, a signal detection analysis of a recollection change account makes clearly different predictions.

Experiment 1

The purpose of Experiment 1 was to test the expectancy heuristic by manipulating relative word frequency between groups during study. MF words were studied either in the context of HF or LF context words and then were tested among MF distractors. To restate our hypothesis, we expected MF targets to receive more remember responses in the HF study context as compared with the LF study context because the MF items would be more memorable than the expected level of memorability in the former condition but weaker than the expected level of memorability in the latter condition.

Method

Subjects. Fifty-six Washington University in St. Louis undergraduates between the ages of 18–23 participated for course credit. Half of the subjects were randomly assigned to the HF study context condition, and half were randomly assigned to the LF study context condition.

Design and materials. Stimulus materials consisted of 36 HF, 36 LF, and 80 MF concrete nouns. The mean hyperspace analog to language (Burgess & Lund, 1997) frequency of the HF, LF, and MF words was 58507, 2504, and 8870, respectively, obtained from the *ellexicon.wustl.edu* Web site (see, Balota et al., in press). All words ranged from four to eight letters, with a maximum of three syllables. All words were concrete nouns, with a minimum concreteness rating of 550 according to the Medical Research Council psycholinguistic database (Wilson, 1988). The ranges of LF, MF, and HF words are consistent with those used previously by other researchers (e.g., Bodner & Masson, 2001; Poirier & Saint-Aubin, 1996). Thirty-six MF words were studied in the context of either HF or LF context words, with four MF buffer words always presented at the beginning of the list and at the end of the study list. The 36 MF targets were tested with 36 MF distractors. The two sets of MF words were counterbalanced across study and test. Thus, 80 words were studied altogether, and there were 72 words on the recognition test. Study and test words were presented in a random order for each subject (with the exception of the buffer words).

The stimuli were presented in the center of the computer screen on IBM-compatible computers with 17-inch (43.18-cm) viewable monitors in 72-point black Times New Roman font. Test words were presented in the same fashion as study words in Arial font.

Procedure. After providing informed consent, subjects were given instructions for the study phase. Subjects were asked to study the words for an upcoming memory test by focusing on the meaning of the words. After the study phase, subjects were given instructions for the recognition test, which took approximately 5 min. Subjects were asked to decide whether they had previously seen the word during the study period. If they had not seen the word, then they were instructed to press a key marked "N," to indicate the word was new. If they had seen the word in the study list, then they were instructed to judge whether the word was recollected, in which case they pressed a key on the keyboard marked with an "R," or if the word was studied but did not include recollective details, then they should press a key marked with a "K" to indicate that they know the word was presented. The difference between remember and know responses was closely based on the instructions by Rajaram (1993) and included examples of each dimension that could be used to provide a remember response and detailed examples of remember and know experiences associated with normal daily activities.

Results

Results of statistical tests were significant at $p < .05$, unless otherwise noted. F values, mean square error, and effect sizes (η^2) are included for each analysis. Table 1 displays the mean hit and false-alarm rate along with the rate of remember-know judgments as a function of study context.

Overall recognition. We begin by examining whether study context affected overall hits and false alarms (i.e., remember + know responses). A one-way analysis of variance (ANOVA) revealed no effect of study context on targets ($F < 1$) or distractors, $F(1, 54) = 1.27$, $MSE = 0.01$, $\eta^2 = 0.02$. Examination of the effect of study context on overall discrimination accuracy (d') also revealed no significant difference ($F < 1$), with values of 2.03 (0.85) and 1.88 (0.64) for the HF and LF contexts, respectively.

Table 1
Percentage of Remember and Know Responses and Overall Hits and False Alarms for Medium-Frequency Targets and Distractors in the High-Frequency and Low-Frequency Study Context Conditions of Experiment 1

Variable	High frequency		Low frequency	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Target				
Remember	48	18	36	20
Know	28	14	36	17
Overall hits	76	18	72	15
Distractor				
Remember	2	3	1	3
Know	14	12	12	7
Overall false alarms	16	13	13	7

Remember-know judgments for targets. Our primary interest was in remember and know responses for studied targets. To reiterate, we expected that MF targets would be given more remember responses in the HF study context than in the LF study context because in the HF study context, MF targets should produce a higher level of memorability than expected relative to the LF study context. Indeed, the hypothesis that study context would influence recollective experience was confirmed. An examination of Table 1 indicates that there were more remember responses for MF targets when they were studied in the context of HF words (48%) than when they were studied in the context of LF words (36%). There were also fewer know responses in the HF context (28%) than in the LF context (36%).

To confirm the apparent difference in remember responses evident in Table 1, we conducted an ANOVA examining the effect of study context (HF, LF) on remember responses, which revealed an increase in remember responses in the HF study context, $F(1, 54) = 5.29$, $MSE = 0.20$, $\eta^2 = 0.09$. The converse was true for know responses, with a significant increase in know responses in the LF study context compared with the HF study context, $F(1, 54) = 4.28$, $MSE = 0.10$, $\eta^2 = 0.07$.

Independence remember-know (IRK) analysis. Yonelinas and colleagues have advocated using remember responses as a direct measure of recollection but with an IRK estimation procedure to compute an estimate of familiarity from know responses (Yonelinas, 2001, 2002; Yonelinas & Jacoby, 1995). They recommend this procedure because the opportunity to make know responses decreases as remember responses increase, assuming the old-new discrimination is fixed and that know responses are made in the absence of recollection. The procedure involves dividing know responses by one minus the rate of remember responses (familiarity = know/[1 - remember]), which essentially conditionalizes know responses on the number of opportunities to make know responses. Although Yonelinas (2001, 2002; Yonelinas & Jacoby, 1995) suggested this procedure because it makes knowing more consistent with the notion that recollection and familiarity operate independently, the present experiments were not designed to examine the issue of independence. However, we include this estimate for completeness and because it seems reasonable to assume that the old-new criterion is fixed (cf. Reder et al., 2000; Rotello,

Macmillan, & Reeder, 2004; Wixted & Stretch, 2004), which would make the proportion of know responses conditional on the rate of remembering. Repeating the study context ANOVA using the IRK estimate of familiarity revealed no effect of study context on studied targets ($F < 1$). Thus, the IRK analysis was not consistent with the direct comparison of the raw know responses and instead suggests that the influence of study context primarily influenced the remember responses.

Remember-know judgments for distractors. We repeated the same analysis as above on the distractors as well. Note that all of the distractors were MF words, just like the targets, and would a priori not be expected to show any effect of the context manipulation. Consistent with this expectation, there was no significant effect of study context on remember or know responses (all $F_s < 1$). However, because remember false-alarm rates were at floor, these data are equivocal with respect to whether they were affected by the context manipulation.

In summary, remember responses were more likely when words were studied with other words of higher relative frequency. These results provide support for the expectancy heuristic, which predicts that studying words in the context of words of higher relative frequency will lead to those words being perceived as more memorable than when they are studied with words of lower relative frequency. Know responses were also affected by study context, with more know responses in the LF context, but analysis using the IRK procedure revealed no effect of study context. Taken together, a change in remember responses with no concomitant change in the hit rate is difficult to explain as the result of changes in memory per se, as exemplified by attention-likelihood theory.

It is important to note that these results are also congruent with previous research manipulating the relative frequency of the study list. Specifically, such manipulations produce very little influence on hit rates (Dorfman & Glanzer, 1988; Malmberg & Murnane, 2002). Our results extend these findings in important ways to suggest that although overall hit rate does not change, there does appear to be systematic changes in the phenomenal experience of remembering. Thus, having subjects report the subjective experience associated with memory retrieval can reveal effects that would not be found by examining the hit rate alone.

Experiment 2

Experiment 1 was successful in demonstrating that study context can influence the use of remember-know responses, without a concomitant influence on hit rates. As mentioned in the introduction, one possible reason for this result is that study context changed the way target words were processed during study (e.g., Glanzer & Adams, 1990). That is, attention may be biased toward items of lower relative frequency during study, such that MF words attract more attention in the HF study context than in the LF study context. This would result in more elaborate encoding of MF words in the HF study context and lead to increased remember responses in this condition. Of course, this attention-likelihood theory (Glanzer & Adams, 1990) also predicts that the hit rate would be affected as well, but one could argue that subtle changes in the allocation of attention during study may affect subjective experience (i.e., remember responses) but not the actual hit rate.

In Experiment 2, relative word frequency was manipulated at test, and, as such, any effects of this manipulation could not be the

result of encoding factors. In Experiment 2, study conditions were held constant across groups, and context was manipulated at retrieval by varying the relative word frequency of the distractors on the recognition test. Subjects studied a homogenous list of MF words and were tested on a subset of them. One group was tested for the MF targets in the context of HF distractors, and the other group was tested on the MF targets in the context of LF distractors. According to the expectancy heuristic, MF targets should receive more remember responses when tested in the HF context than in the LF context, based on differences in the expected memorability of the test items (see Benjamin, 2003; Guttentag & Carroll, 1998).

Method

Subjects. Fifty-six Washington University in St Louis undergraduates between the ages of 18–23 years participated in the experiment for course credit.

Materials. The word pool for Experiment 2 was the same as the one used in Experiment 1. For both the HF and LF test context groups, 36 MF target words were studied in the context of the 44 MF filler words, including four buffers at the beginning and four buffers at the end of the study list. The distractors on the recognition test were the 36 HF or 36 LF words that had been used as study context words in Experiment 1. That is, in the HF test context group, the MF targets were tested with 36 HF distractors, whereas in the LF test context group, the MF targets were presented with 36 LF distractors. Thus, all subjects received the same homogenous list of 80 MF words during study. The groups differed in that one group was tested for the 36 MF targets with 36 HF distractors, and the other group was tested for the MF targets with 36 LF distractors.

Results

The hit and false-alarm rate along with the rate of remember and know responses as a function of word frequency context at test are displayed in Table 2.

Overall recognition. As shown in Table 2, the overall hit rate for targets (remember + know) was numerically greater for the HF than for the LF context (.78 and .72, respectively), though this difference did not reach the conventional level of significance,

Table 2
Percentage of Remember and Know Responses and Overall Hits and False Alarms for Medium-Frequency Targets and Distractors in the High-Frequency and Low-Frequency Test Context Conditions of Experiment 2

Variable	High frequency		Low frequency	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Target				
Remember	49	20	39	14
Know	30	12	33	12
Overall hits	78	14	72	11
Distractor				
Remember	2	4	2	4
Know	19	13	8	6
Overall false alarms	21	15	9	9

$F(1, 54) = 3.34$, $MSE = 0.03$, $\eta^2 = 0.06$, $p = .073$. The overall false-alarm rate (remember + know) was numerically greater for the HF than for the LF test context as well (.21 and .09, respectively), and this difference was statistically significant, $F(1, 54) = 13.39$, $MSE = 0.10$, $\eta^2 = 0.20$. Examination of the effect of test context on overall discrimination accuracy (d') revealed a nonsignificant difference, $F(1, 54) = 2.37$, $MSE = 1.13$, $\eta^2 = 0.04$, with means of 1.82 ($SD = 0.81$) and 2.10 ($SD = 0.54$) for the HF and LF contexts, respectively. Given that the overall hit and false-alarm rates were somewhat higher in the HF context, one might conclude that responding was simply more liberal in the HF context than in the LF context. However, as discussed below, the higher hit rate in the HF context was the result of increases in remember responses whereas the higher false-alarm rate in the HF context was the result of increased know responses, which is inconsistent with the most straightforward signal detection account (e.g., Donaldson, 1996). Of course, given that the distractors were of different frequencies across groups, this finding may be artifactual.

Remember-know judgments for targets. As in the previous experiments, our primary interest in Experiment 2 were remember-know responses for studied targets. This experiment differed, though, in that we were examining whether relative word frequency at the time of test influenced remember responses, which would be confirmed if MF targets were given more remember responses in the HF test context (i.e., with HF distractors) than in the LF test context (i.e., with LF distractors). Indeed, as shown in Table 2, there were more remember responses in the HF test context (49%) than in the LF test context (39%), consistent with the findings of study context in Experiment 1.

To confirm the apparent difference in remember responses across groups, we examined remember responses as a function of test context (HF, LF), which revealed a significant increase in remember responses in the HF test context, $F(1, 54) = 4.27$, $MSE = 0.12$, $\eta^2 = 0.07$. There was no effect of test context on know responses ($F < 1$). Thus, test context primarily appears to have affected remember responses exclusively, rather than affecting both remember and know responses.

IRK analysis. Repeating the study context ANOVA on studied targets using the IRK estimate of familiarity revealed no effect of study context ($F < 1$). Hence, although remember responses are changing as a function test context, there is no disproportionate change in know responses, replicating the IRK results from Experiment 1.

Remember-know judgments for distractors. Note that in Experiment 2, distractors were HF words in the HF test context and LF words in the LF test context. Thus, a priori, we would expect higher rates of know responses for HF distractors, consistent with prior research (Balota, Burgess, Cortese, & Adams, 2002; Gardiner & Java, 1990; Guttentag & Carroll, 1997; Joordens & Hockley, 2000; Reder et al., 2000). As one might expect, remember responses for distractors were quite low for both test context groups (< 2%) and did not differ ($F < 1$). By contrast, know responses were higher for the HF test context compared with the LF test context, $F(1, 54) = 16.84$, $MSE = 0.19$, $\eta^2 = 0.24$. However, as with Experiment 1, it is not clear whether context affected the remember false-alarm rate because these responses were at floor, and, thus, these data are equivocal with respect to whether they were affected by the context manipulation.

In summary, Experiment 2 replicated the finding that remember responses increased in the HF context, though in this case the context manipulation occurred during the test phase. This result favors a decision-based mechanism, such as the expectancy heuristic, rather than a purely encoding-based explanation. Although it is possible that the study context manipulation may have biased attention at encoding in Experiment 1 or may have led to differences in list strength across study context conditions, this could not have produced the obtained differences in remember responses in Experiment 2. We propose a decision-based mechanism to explain the data we report, such that differences in the expected versus actual level of memorability influenced remember-know decisions across both study and test context conditions. However, the present data cannot be used to totally rule out all memory-based explanations.

Experiment 3

The purpose of Experiment 3 was to more directly compare the expectancy heuristic and recollection change accounts via a manipulation of memory strength and an explicit manipulation of subjects' expectancies. Subjects studied a mixture of strong (studied five times) and weak (studied once) items but were only tested on the weak items. In the expecting weak group, subjects were told that they were only being tested on the weak items, but in the expecting strong group, subjects were not told this and so they were expecting both strong and weak items. Thus, the overall expected level of strength of the studied targets was considerably greater in the expecting strong group. As a result, subjects in the expecting weak group were expected to give more remember responses to studied words. Moreover, if remember responses for distractors were above the floor effects seen in the relative word frequency experiments, then this effect would also be expected to occur for distractors as well. It is important to note that if these expectancies operate on decision processes rather than by changing the type or amount of information retrieved, then greater use of remember responses in the expecting weak group should not affect discrimination accuracy. In contrast, if the expecting strong group changes the quality of the memory retrieved compared with the expecting weak group, then one might expect a change in memory discrimination.

Because the expectancy heuristic predicts changes in the criterion for remember responses but a memory-based mechanism predicts changes in discrimination (e.g., d'), signal detection analysis can be used to compare these explanations (cf. McCabe & Smith, 2006). A basic equal-variance signal detection model of remember-know is shown in Figure 2 (top panel) for illustrative purposes. According to this signal detection model, subjects adopt two response criteria on a remember-know test, with the remember criterion having a high threshold and the know response having a more liberal threshold that distinguishes targets and lures. If this model is used as an illustration of subjects from the expecting weak group in Experiment 3, then the bottom panel illustrates the prediction of the expectancy heuristic for subjects in the expecting strong group. The dotted arrow shows the conservative shift of the remember criterion, such that subject's would require more information from the study episode before endorsing an item as remembered. An important prediction of this signal detection instantiation of the expectancy heuristic is that although the remember criterion

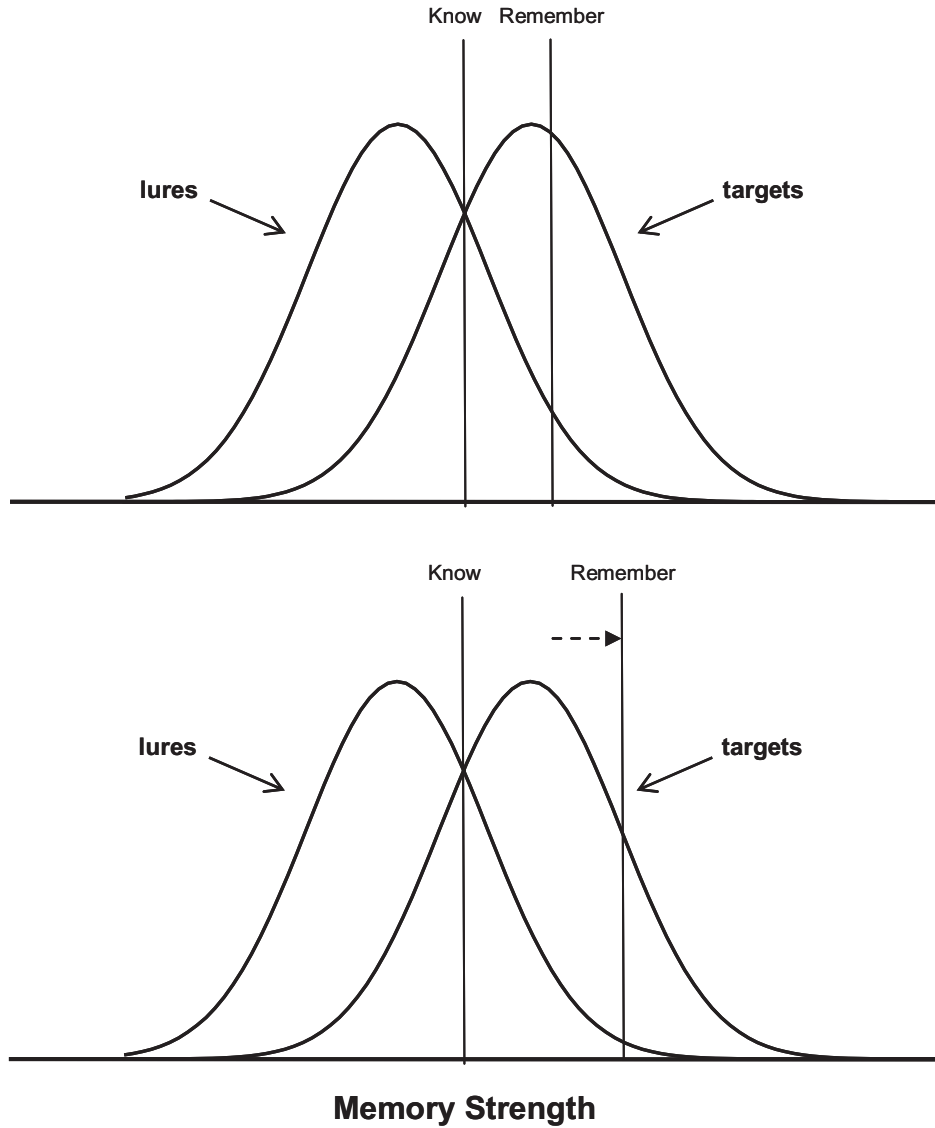


Figure 2. Signal detection model predictions for the expectancy heuristic explanations for Experiment 3. The top panel shows the equal-variance remember–know model representing the expecting weak group in Experiment 3. The bottom panel shows the prediction of the expectancy heuristic model for the expecting strong manipulation.

changes, the know criterion would not necessarily change but know responses should increase in the group expecting stronger or more distinct items. Specifically, as the remember criterion moves to the right, remember responses will decrease, but if the know criterion does not change, then know responses will concomitantly increase (because hit rate is the sum of remember and know responses for old items). Of importance, if the level of remember false alarms is above the floor levels found in the first two experiments, then one would not only see this effect on targets but on lures as well. Thus, the net effect of these changes, a shift in the remember criterion with no change in the know criterion, should not influence discriminability (d'). This can be seen by comparing the target and lure distributions in the top and bottom panels, which are equidistant.

A signal detection instantiation of the recollection change model prediction for the expecting strong group is shown in Figure 3. The recollection change model would be supported if subjects retrieved less information in the context of more distinct or more memorable items. According to this explanation, remember responses are reduced for targets because the target and lure distributions move closer to one another. The dashed arrow shows the movement of the target distribution in the bottom panel relative to the expecting weak group in the top panel. It is important to note that although remember responses to targets would be reduced in the expecting strong group, remember responses to distractors would not, resulting in an overall decrease in the accuracy of remember responses. Thus, the recollection change model makes the prediction that d' for remember responses should be reduced in the expecting strong

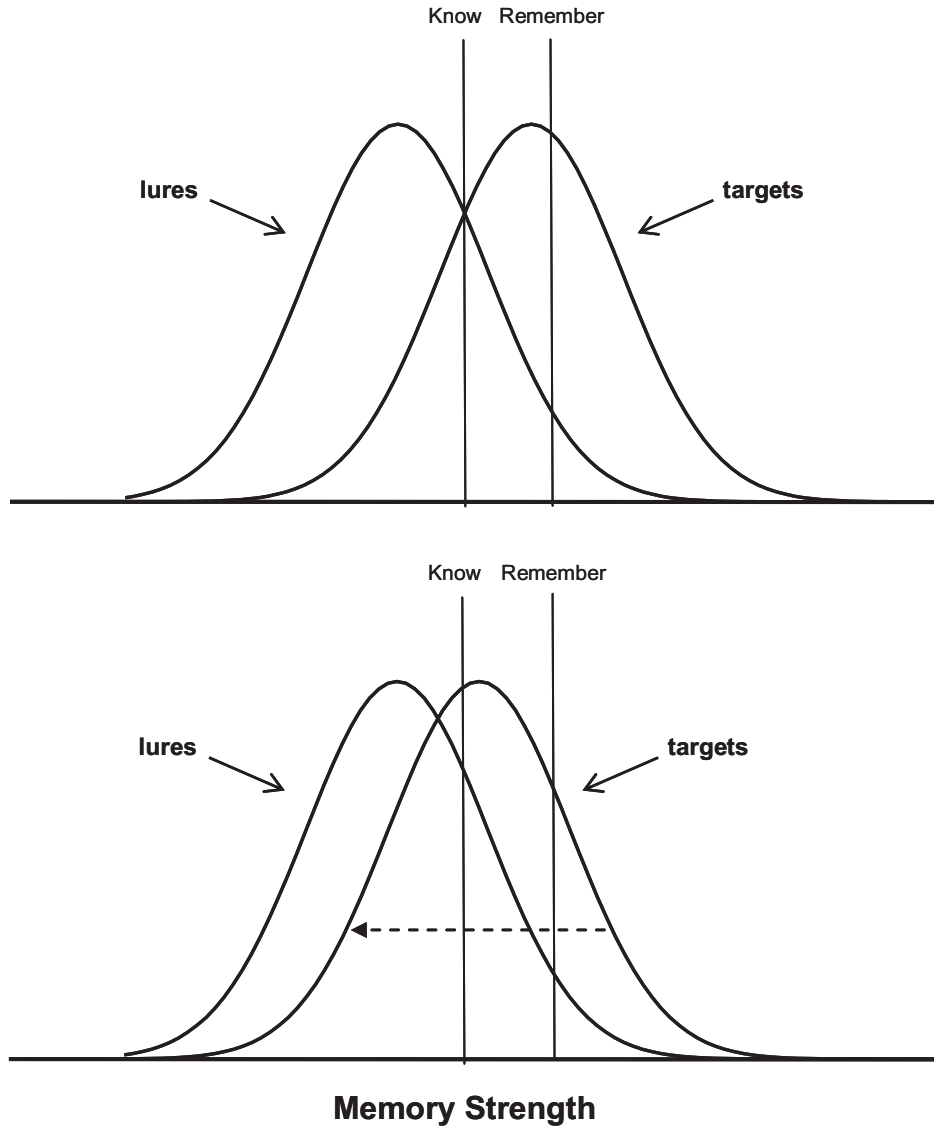


Figure 3. Signal detection model predictions for the recollection change model for Experiment 3. The top panel shows the equal-variance remember–know model representing the expecting weak group in Experiment 3. The bottom panel shows the prediction of the recollection change model for the expecting strong manipulation.

group, and, consequently, overall discrimination accuracy (overall d') should also be reduced.

Method

Subjects. Fifty-six Colorado State University undergraduates between the ages of 18–23 participated for course credit. Half of the subjects were randomly assigned to the expecting strong group, and half were randomly assigned to the expecting weak group.

Design and materials. Stimulus materials consisted of 120 words with a frequency ranging from medium to high (based on the values in Experiments 1 and 2) and a concreteness rating between 400–550, according to the Medical Research Council psycholinguistic database. We assumed that using less concrete

words would increase the false-alarm rate to distractors, which would allow a less ambiguous test of the expectancy heuristic than in Experiment 2. Words were divided into two sets of 60, with words randomly assigned to each set and one set serving as the studied items and the other set serving as the distractors. Of the 60 studied words, 30 were studied one time each and 30 were studied five times each, randomly intermixed on the study list. The two sets of words were counterbalanced across subjects, serving as the studied items and distractors an equal number of times.

Procedure. The procedure followed that of Experiments 1 and 2, except that subjects were informed before study that they would be seeing some of the items once and others five times. In addition,

the expecting weak group was given the following instruction right before taking the recognition test:

You studied some of the words once and other words five times. None of the words you studied five times are going to be on the test. Only words you studied once will be on the test. So, you will either have seen the word once during the study phase, or it will be a new word that you had not seen during study.

Results

The hit and false-alarm rate along with the rate of remember and know responses and signal detection measures are displayed in Table 3.

Hits and false alarms. We began by examining whether instruction context affected overall hits and false alarms (i.e., remember + know responses). A one-way ANOVA revealed no effect of study context on targets or distractors ($F_s < 1$).

Remember-know judgments for targets. As shown in Figure 4, there were more remember responses in the expecting weak context (47%) than in the expecting strong context (35%), $F(1, 54) = 4.86, MSE = 0.21, \eta^2 = 0.08$. There were also fewer know responses in the expecting weak context (24%) than in the expecting strong context (33%), though this difference did not reach conventional significance levels, $F(1, 54) = 3.24, MSE = 0.11, \eta^2 = 0.06$.

IRK analysis. Repeating the instruction context ANOVA on the IRK estimate of familiarity for studied targets revealed no effect of instruction context ($F < 1$).

Remember-know judgments for distractors. The expectancy heuristic predicts that a pattern similar to that shown for targets should occur for distractors, whereas the recollection change account does not. As seen in Figure 4, the pattern of results for distractors was similar to that of targets. Specifically, there were

more remember responses in the expecting weak context (10%) than in the expecting strong context (3%), $F(1, 54) = 5.96, MSE = 0.06, \eta^2 = 0.10$. There were also fewer know responses in the expecting weak context (15%) than in the expecting strong context (19%), though this difference was not significant, $F(1, 54) = 1.72, MSE = 0.02, \eta^2 = 0.03$. Thus, in terms of the statistical patterns of remember-know responses, the effect of the instruction manipulation was identical for targets and distractors, supporting the predictions of the expectancy heuristic.

Signal detection analyses: Discrimination (d') and bias (β). As outlined in the introduction, the recollection change and expectancy heuristic accounts make differing predictions with respect to signal detection models of remember-know judgments in Experiment 3. The recollection change account predicts differences in discrimination accuracy (e.g., d') but not necessarily in decision criteria, and the expectancy heuristic predicts differences in the remember criterion (e.g., remember β) but no concomitant change in discrimination accuracy.

Overall, there were no changes in discrimination accuracy for remember d' or overall d' ($F_s < 1$), which is consistent with the expectancy heuristic model. There were no changes in discrimination accuracy for know d' either, $F(1, 54) = 1.56, MSE = 0.35, \eta^2 = 0.03$.

Given no differences between the instruction groups on overall d' , the critical test of the expectancy heuristic is whether there were changes in remember β . Indeed, subjects in the expecting strong group were more conservative in their use of the remember criterion as compared with subjects in the expecting weak group, $F(1, 54) = 5.25, MSE = 3.88, \eta^2 = 0.09$. There were no differences between groups in overall β ($F < 1$) or know β , $F(1, 54) = 2.22, MSE = 0.64, \eta^2 = 0.04$.

The equivalence of the hit and false-alarm rates between the expectancy instruction groups, and consequently d' as well, indicates that the instructions did not affect the amount or type of information that was retrieved. Nonetheless, expectancies about the strength of studied items affected the remember criterion. These data support the operation of an expectancy heuristic, a decision-based mechanism, in influencing remember-know responses. When subjects were expecting items to be stronger than they actually were, they required more studied information to be retrieved before they would call an item remembered. Although the expected pattern in know responses was consistent with the expectancy heuristic as well, with decreases in knowing in the expecting weak as compared with expecting strong groups, these differences did not reach levels of statistical significance. This may be due to any of several factors, but it is partly the result of a slightly more liberal overall response bias in the expecting weak group, who gave 3% more hits and false alarms compared with the expecting strong group. Thus, although the effect of instructions on knowing did not reach statistical significance, this effect was diminished by a somewhat more liberal responses bias in the expecting weak group. It is also worth noting that the findings reported here may not necessarily generalize to all instructions used to collect remember-know responses. For example, some instructions discourage remember false alarms by requiring subjects to justify their remember responses (Yonelinas, 2001). If these types of instructions are successful at keeping remember false alarms at floor, then it would not be possible to assess whether the remember criterion or the recollection process has

Table 3
Percentage of Remember and Know Responses, Overall Hits and False Alarms, and Signal Detection Estimates for the Expecting Weak and Expecting Strong Groups in Experiment 3

Variable	Expecting weak		Expecting strong	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Target				
Remember	47	22	35	19
Know	24	18	33	18
Overall hits	71	16	68	15
Distractor				
Remember	10	13	3	5
Know	15	10	19	12
Overall false alarms	25	17	22	14
Signal detection estimate: d'				
Remember	1.49	0.65	1.52	0.58
Know	0.33	0.50	0.48	0.45
Overall	1.44	0.54	1.37	0.57
Signal detection estimate: β				
Remember	1.25	0.93	1.77	0.78
Know	0.14	0.61	0.35	0.45
Overall	0.15	0.83	0.30	0.61

Note. d' is an estimate of signal detection discrimination, and β is an estimate of signal detection bias.

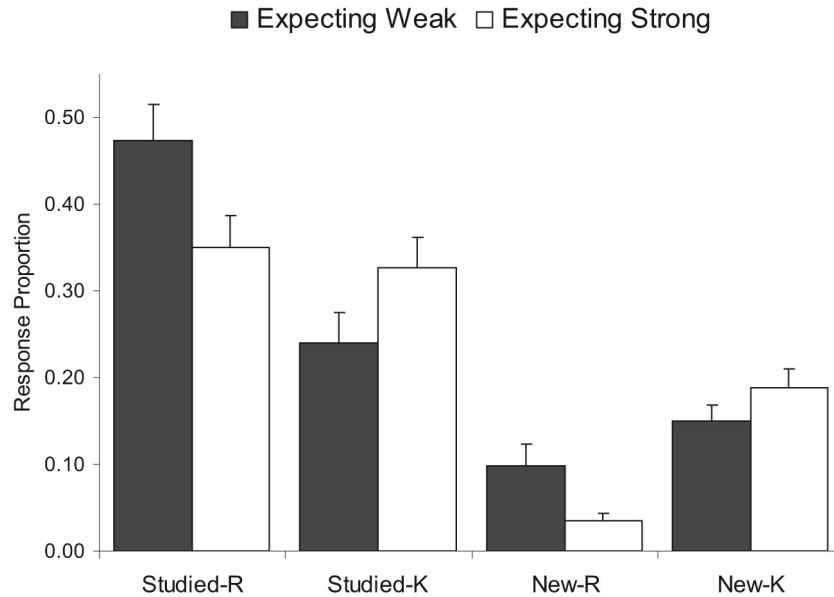


Figure 4. Remember (R) and know (K) responses for studied and new items for the expecting weak and expecting strong conditions in Experiment 3. Error bars present the standard error of the mean.

been affected. Our use of low concreteness words and the memory strength and instructional manipulations was intended to encourage remember false alarms so as to allow a comparison of the expectancy heuristic and recollection change accounts, but more typical situations might not lend themselves to these sorts of comparisons.

General Discussion

To summarize the results, the influence of relative word frequency on remember–know judgments in Experiments 1 and 2 revealed that encountering MF target words in the context of HF words resulted in those targets receiving more remember responses than when those same words were encountered in the context of lower frequency words. This effect occurred regardless of whether context was manipulated at study, when studied items differed in relative frequency, or at test, when distractors differed in their relative frequency. We proposed an expectancy heuristic to explain these findings, such that remember responses were affected by the expected level of distinctiveness of items in a given context. In contexts in which subjects expected distinctive items on the test, they were less likely to give remember responses to studied words than when they expected less distinct items. Because explanations based on the notion that context was affecting the amount or type of information retrieved could not be totally ruled out in Experiments 1 and 2, Experiment 3 was conducted to more directly adjudicate between decision-based and memory-based accounts. In Experiment 3, subjects who were expecting remembered items to include more recollective details were less likely to use remember responses and more likely to use know responses for both studied targets and distractors than were subjects expecting less detail to be retrieved for studied items.

The Expectancy Heuristic and Extant Systems and Processing Accounts of Remember–Know Judgments

As we mentioned at the outset, the remember–know procedure was originally developed to examine whether people could differentiate between different states of awareness associated with memory retrieval, with these states of awareness presumably being associated with different memory systems (e.g., episodic and semantic). However, it is unclear how or why context might affect the output of different memory systems. Why would expectancies about the memorability of studied items lead to an increased or decreased reliance on the episodic memory system? Although our data do not rule out the notion that remember and know responses are associated with different memory systems, nor were they intended to, they do suggest that there are other important influences on remember–know responses beyond the output of different memory systems.

Similarly, our data do not necessarily cast doubt on the notion that remember responses are associated with a recollection process and know responses are associated with a familiarity process, but our data do suggest remember and know judgments are influenced by more than just the recollection and familiarity processes, respectively. If remember responses did reflect a threshold recollection process, then why would this recollection process be disrupted by changes in the memorability of study or test context, including changes in expectancies at the time of test? This might occur if the context in which items were studied and/or tested changed the way in which items were processed (cf. Jacoby, Shimizu, Velanova, & Rhodes, 2005), but a change in recollection would typically be expected to change the overall hit rate as well. Of course, development of systems and processing theories have traditionally been aimed at explaining various experimental dissociations rather than explaining the decision processes involved in subjective experi-

ence of remembering and knowing. The present study focuses more on the latter decision processes.

In this light, our results are not necessarily inconsistent with dual-process models, these models simply need to be supplemented by an explanation of how decision processes might affect the experience of remembering and knowing. That is, it is possible for context to affect remember and know responses independent of recollection and familiarity processes. The important implication of this finding, though, is that remembering and knowing should not be interpreted as representing reliable estimates of cognitive processes (or systems for that matter) but rather should be treated as subjective memory judgments that can be affected by changes in the underlying memory processes or by influences on monitoring processes. Thus, influencing the recollection and familiarity processes would be expected to change remember-know judgments, but contextual manipulations can also affect how information is weighted when making memory decisions. Note, too, that just because we showed that context affected decision processes in the current experiments does not rule out the possibility that context can also influence the type and amount of information retrieved (see Bodner & Lindsay, 2003).

The functional account of recollection includes aspects of traditional dual-process memory theories and also provides an explanation for how contextual factors influence recollection (Bodner & Lindsay, 2003; Gruppuso et al., 1997). According to the functional account, the subjective experience of recollection depends on the context in which remembering occurs, with remember judgments occurring when retrieved information allows subjects to complete a criterial memory test. By this account, retrieval of the same information may be experienced as recollection or familiarity, depending on the context in which the judgment is made. Our expectancy heuristic is consistent with this idea but is perhaps a bit broader in its purported scope. According to the expectancy heuristic, the subjective experience of remembering is defined by a discrepancy from some average level of memorability rather than being directly tied to successful completion of a criterial task. We should reiterate that our expectancy heuristic and the functional account are similar in that both assume that remember judgments accurately reflect the conscious experience of recollection and that this experience of recollection differs depending on the context in which those judgments are made. This is a crucial point of agreement because this issue is a point of some debate in the literature, with others having suggested that remember-know responses do not accurately reflect discrete states of awareness associated with memory retrieval (e.g., Donaldson, 1996; Dunn, 2004).

The distinctiveness-fluency framework (Rajaram, 1998; Rajaram & Geraci, 2000) is compatible, in principle, with our results as well. For this framework to account for the present context effects, one must assume that the perception of distinctiveness occurs when items exceed some prespecified level of memorability whereas a perception of fluency occurs when items are familiar but do not exceed this level of memorability. Thus, the expectancy heuristic offers a plausible mechanism to explain how subjects come to define items as distinctive. In general, items are said to be distinct, or salient, if they are unusual relative to some background context. Indeed, our context effects can be thought of as just this sort of situation. In the study context situation, MF items are unusually memorable relative to the average memorability of the items in the HF, as compared with LF, context. Thus, if distinctiveness

is defined as relative memorability, then distinctiveness certainly differed as a function of study context in Experiment 1. In the test context situation, if the expected level of distinctiveness was influenced by the characteristics of the distractors, then one would again expect the MF items to seem unusually memorable relative to the HF, as compared with the LF, distractors. Finally, in Experiment 3, when subjects were only expecting once-presented items the test items would seem more memorable compared with a situation in which subjects were expecting to encounter studied items that had been presented once or five times. Thus, expectations about the memorability of items leads subjects to define distinctiveness (as reflected by the remember judgments) differently.

Although we have attempted to outline a framework for understanding how subjects construct remember-know judgments online as the result of retrieval expectations, we note as well that there are plausible memory-based explanations for the relative word frequency data from Experiments 1 and 2. For example, a cuing explanation based on possible differences in the cuing properties of LF and HF words could explain how relative word frequency could affect the amount or type of information studied or retrieved in the HF and LF contexts. This explanation is similar to attention-likelihood theory in that it suggests that the MF targets were processed differently in the HF and LF word contexts and provides a plausible explanation for our study context manipulation. However, can this same explanation account for the test context manipulation? According to the cuing explanation, HF distractors may have been better cues for the MF targets than were the LF distractors. This is because HF words were more likely to be preexperimentally associated with other words than were LF words, thus the HF distractors should have been more likely to act as effective retrieval cues for targets on the recognition test. This explanation cannot be ruled out on the basis of the present data; however, there are published data that question that such powerful cuing effects would occur as the result of differences in word frequency. Specifically, although there are differences in the number of contexts in which HF and LF words occur, data from free association norms show that the direction of the associations is such that higher frequency words have more connections coming from other words than do lower frequency words, but there are no differences in the number of connections going to other words (Nelson & McEvoy, 2000). This latter finding casts doubt on this cuing explanation of both the study and test contexts, as this explanation would require that HF words have a greater number of connections going to other words (specifically, MF words in Experiment 2). In addition, and perhaps more importantly, empirical findings show that LF and HF words do not differ in their effectiveness as retrieval cues at test, presumably because of the equivalent number of connections going to other words (Nelson & McEvoy, 2000). Because the associative memory processes driving remember judgments and cued recall are very similar, this finding casts doubt on this cuing explanation, particularly because subjects in Experiment 2 were presumably not even trying to use the distractors to cue memory of targets. Nonetheless, the cuing explanation that we outlined here cannot be ruled out completely by the present data, and examining alternative explanations will require further research.

Partly because it was not possible to unequivocally determine the mechanism causing the relative word frequency effects in

Experiments 1 and 2, Experiment 3 was conducted to directly examine whether the expectancy heuristic actually operates on expectancies. These data provided strong support for the notion that remember-know responses can be affected by the expected strength of test items. Because the study episode and the tests were identical for all subjects in Experiment 3, with only expectancies about the strength of the test items being manipulated, the effects observed could not be due to some cuing properties of the stimuli. Instead, it appears that subjects adopted a level of expected memory strength and defined remembering relative to that level of expected strength.

The signal detection models provided an important test between the expectancy heuristic and an alternative memory-based explanation that we referred to as the recollection change account. Specifically, the expectancy heuristic predicts a shift in the remember criterion but no change in discrimination accuracy. That is, it should be possible to move the remember criterion (β) selectively for items of the same memory strength, thereby changing the proportion of remember-know responses for targets and lures without changing d' . Alternatively, contextual manipulations could also influence the type or amount of information retrieved, which would be expected to change discrimination accuracy, perhaps selectively for remember responses, rather than affecting the remember criterion. Thus, the recollection change model predicts changes in overall d' , or remember d' , but no change in β . In fact, the data from Experiment 3, where subjects were expecting weak or strong items on the test, showed selective effects on the remember criterion (β) without any concomitant changes in overall d' or remember d' . Thus, the expectancy manipulation in Experiment 3 provided strong converging support for the expectancy heuristic rather than alternative memory-based accounts.

We note, too, that the expectancy heuristic is compatible with several extant signal detection models of remember-know judgments. In particular, our data indicate that decision processes can be influenced by the context in which remember-know judgments are made. However, we should also note that just because our data and theory are compatible with signal detection models does not mean we advocate a single-process model of recognition memory. Indeed, several signal detection models that are consistent with dual-process accounts of memory have recently been proposed (Banks, 2000; Rotello et al., 2004; Wixted & Stretch, 2004). These models are based on the notion that multiple processes (or types of evidence) contribute to recognition decisions, including remember-know responses. For example, Wixted and Stretch (2004) have suggested that it makes little sense to base responding on one process or another (e.g., recollection or familiarity), but rather, even if there are two processes operating to influence successful retrieval, subjects likely sum these processes when making their recognition decisions. Rotello et al. (2004) have gone further, suggesting that the remember criterion is chosen on the basis of the difference between global and specific strength of the items studied. Our expectancy heuristic is compatible with this idea in that we propose that subjects use some average level of memorability (based on recollection and familiarity processes) and use that level to set their decision criterion for what constitutes a remember response. It is worth reiterating, though, that unlike some signal detection theorists (e.g., Donaldson, 1996; Dunn, 2004), we believe that remember-know judgments reflect different experiential states associated with memory retrieval, which is

not necessarily incompatible with the notion that memory strength is continuous.

Summary and Conclusions

The decision-based model we call the expectancy heuristic provides a compelling account of current and past context effects on remember-know judgments. This study also suggests that focusing attention on the decision processes that influence remember-know judgments is required to gain a comprehensive understanding of the subjective experience associated with memory retrieval.

References

- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). Word-frequency mirror effect in young, old, and early stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*, 199–226.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (in press). The English lexicon project. *Behavior Research and Methods*.
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*, 267–273.
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, *31*, 297–305.
- Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563–580.
- Bodner, G. E., & Masson, M. E. J. (2001). Prime validity affects masked repetition priming: Evidence for an episodic resource account of priming. *Journal of Memory and Language*, *45*, 616–647.
- Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology*, *29*, 461–473.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, *12*, 177–210.
- Davis, J. C., Lockhart, R. S., & Thomson, D. M. (1972). Repetition and context effects in recognition memory. *Journal of Experimental Psychology*, *92*, 96–102.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523–533.
- Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language*, *27*, 633–648.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, *111*, 524–542.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.
- Gardiner, J. M., Gawlick, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering: Elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*, 107–110.
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23–30.
- Gardiner, J. M., Kaminska, Z., Dixon, M., & Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review*, *3*, 366–371.
- Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review*, *13*, 466–473.
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition

- memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 5–16.
- Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 259–278.
- Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*, 37, 502–516.
- Guttentag, R. E., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory & Cognition*, 26, 951–958.
- Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word frequency and word concreteness. *Journal of Experimental Psychology: Learning Memory and Cognition*, 23, 1306–1323.
- Huron, C., Giersch, A., & Danion, J. M. (2002). Lorazepam, sedation, and conscious recollection: A dose–response study with healthy volunteers. *International Clinical Psychopharmacology*, 17, 19–26.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306–340.
- Jacoby, L. L., Kelley, C., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.
- Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*, 52, 493–504.
- Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1534–1555.
- Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 616–630.
- McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review*, 11, 1074–1079.
- McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory*, 14, 570–583.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533.
- Nelson, D. L., & McEvoy, C. L. (2000). What is this thing called frequency? *Memory & Cognition*, 28, 509–522.
- Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 1083–1094.
- Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness and Cognition*, 4, 387–398.
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, 50, 408–412.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21, 89–102.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 365–377.
- Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review*, 5, 71–78.
- Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1070–1074.
- Reder, L. M., Nhouyvasivong, 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, and Cognition*, 26, 294–320.
- Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, 111, 588–616.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460–471.
- Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognition and presuppositional strategies. *Journal of Memory and Language*, 33, 203–217.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.
- Westerman, D. L., & Greene, R. L. (1998). The revelation that the revelation effect is not due to revelation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 377–386.
- Westerman, D. L., Lloyd, M. L., & Miller, J. M. (2002). On the attribution of perceptual fluency in recognition memory: The role of expectation. *Journal of Memory and Language*, 47, 607–617.
- Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autoegetic consciousness. *Psychological Bulletin*, 121, 331–354.
- Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 547–565.
- Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, Version 2. *Behavioral Research Methods, Instruments and Computers*, 20, 6–11.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, 11, 616–641.
- Yonelinas, A. P. (2001). Consciousness, control and confidence: The three Cs of recognition memory. *Journal of Experimental Psychology: General*, 130, 361–379.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*, 34, 622–643.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember/know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, 12, 1–17.

Received June 15, 2006

Revision received January 17, 2007

Accepted January 17, 2007 ■