Consciousness and Cognition 24 (2014) 57–69

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Investigating the subjective reports of rejection processes in the word frequency mirror effect

J. Thadeus Meeks [a,*], Justin B. Knight [b], Gene A. Brewer [c], Gabriel I. Cook [d], Richard L. Marsh [e]

[a] Department of Psychology, Southern Illinois University Edwardsville, United States
[b] BioImaging Research Center & Department of Psychology, University of Georgia, United States
[c] Department of Psychology, Arizona State University, United States
[d] Department of Psychology, Claremont McKenna College, United States
[e] Department of Psychology, University of Georgia, United States

ARTICLE INFO

Article history:
Received 15 May 2013

Keywords:
Recognition memory
Word frequency
Rejection processes
Remember–Know

ABSTRACT

We sought to systematically investigate how participants subjectively classify the basis of their recognition memory judgments for low and high word frequency items. We found that participants more often reported rejection processes related to the increased perceived memorability for unstudied low word frequency items (relative to high word frequency items), rather than classifying their decision on a lack of familiarity. Experiment 2 replicated this pattern and demonstrated context variability and word frequency independently influenced the subjective classifications for correct rejections. Results of Experiment 3 revealed that these differences are dependent upon having experience with both low and high frequency items. Overall, these data suggest participants' rejection of low frequency items is more strongly related to judgments of perceived memorability, but only when they are presented in the context of high frequency items. The results are discussed in relation to distinctiveness and expected memorability.

1. Introduction

The investigation of determining whether an item has occurred in a particular context has a long and venerable history in the experimental psychology literature (Strong, 1912; for recent reviews see Malmberg, 2008 and Yonelinas, 2002). One regularity emerging from this literature is the occurrence of mirror effects. Mirror effects appear when memory for a particular class of items exhibits higher hit rates and lower false alarm rates than for a differing class of items (for a review, see Glanzer, Adams, Iverson, & Kim, 1993). There are many types of mirror effects in recognition memory, but the most commonly studied effect is related to normative word frequency (WF; e.g., Glanzer & Adams, 1985). Words with low word frequency (LWF; e.g., *aardvark*) tend to be correctly recognized and false alarmed to less than words with high word frequency (HWF; e.g., *city*). Many proposals have been made about what causes the mirror effect; however, systematic investigation of the potential processes involved in the effect has primarily focused on the increased hit rate for LWF items. As with the study of mirror effects, the predominant approach to studying recognition memory has been concerned with how individuals correctly identify old items with relatively little attention paid to how individuals reject new items. Our focus in this study was on the possible contributions of rejection processes involved in the increased correct rejection rate (i.e., lower false alarm rate) for LWF items in traditional mirror effects. More specifically, we wished to investigate the subjective reports of memory processes associated with the decision that LWF and HWF items have not been studied (and the report of rejection processes in general).

* Corresponding author.
E-mail address: jmeeks@siue.edu (J.T. Meeks).

1053-8100/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.concog.2013.12.007

2. The hit rate portion of the word frequency mirror effect

The hit rate portion of the WF mirror effect is often attributed to enhanced recollection for LWF items (thus higher hit rates), whereas the false alarm portion is often attributed to greater pre-experimental familiarity for HWF items (e.g., Arndt & Reder, 2002; Guttentag & Carroll, 1994; Joordens & Hockley, 2000; Reder et al., 2000; Rugg, Cox, Doyle, & Wells, 1995). Some evidence for the role of recollection in the hit rate portion of the WF mirror effect comes from examining participants' subjective classification of their old responses (Cook, Marsh, & Hicks, 2006; Dewhurst, Brandt, & Sharp, 2004; Dewhurst, Hitch, & Barry, 1998; Gardiner & Java, 1990; Joordens & Hockley, 2000; Reder et al., 2000). In the Remember–Know procedure, "Remember" responses represent conscious recollection of the episodic context associated with the studied item. Alternatively, if the item is familiar enough to warrant an old response but lacks any diagnostic episodic detail, the participants should respond "Know". LWF items are given more Remember responses than HWF items (Cook et al., 2006; Dewhurst et al., 1998, 2004; Gardiner & Java, 1990; Joordens & Hockley, 2000).

There is some debate about the use of subjective reports in measuring recollection and familiarity. For example, many claim that the binary judgment between Remember and Know simply reflects different confidence levels in the recognition decision (Donaldson, 1996; Dunn, 2004; Wais, Mickes, & Wixted, 2008; but see Williams, Conway, & Moulin, 2013 for recent evidence against this). There is also evidence that the amount of Remember and Know responses is altered by list composition at study (e.g., McCabe & Balota, 2007) and test (e.g., Benjamin, 2005; Bodner & Lindsay, 2003; McCabe & Balota, 2007; Tousignant & Bodner, 2012). Despite these limitations, if used correctly, the Remember–Know procedure can be used to successfully measure recollection and familiarity (see Migo, Mayes, & Montaldi, 2012 for a recent review and Lampinen, Neuschatz, & Payne, 1997 for further discussion of the pros and cons of subjective memory reports including Remember–Know judgments).

3. The correct rejection portion of the word frequency mirror effect

While most studies have focused on the hit rate portion of the WF mirror effect, much less research has examined the false alarm portion. The aim of the current study was to assess the cognitive mechanisms that contribute to the false alarm portion of the WF mirror effect by utilizing subjective memory classifications similar to those requested in the Remember–Know procedure. Another way to conceptualize the false alarm portion of the WF mirror effect is that new LWF items are more often correctly rejected than HWF items. Because LWF items lack pre-experimental familiarity relative to HWF items, they elicit fewer implicit associative responses (Underwood & Freund, 1970). This lower level of familiarity may result in greater rejections of new LWF items as studied in the experimental context. Another possible way to explain the correct rejection differences focuses on the metacognitive assessment of LWF items. Presumably, due to their low rate of pre-experimental occurrence, LWF items are judged to be distinctive or memorable enough to lead to the metacognitive impression that they would have been remembered if they had in fact been previously experienced (Benjamin, 2003; Brown, Lewis, & Monk, 1977; Guttentag & Carroll, 1998). This in turn would lead to a higher correct rejection rate for LWF items than HWF items. This metacognitive, memorability-based judgment requires that the participant realize that LWF items are more memorable than HWF items. Some, however, have found that participants actually judge high frequency words as more memorable (e.g., Wixted, 1992). Guttentag and Carroll, as well as Benjamin, asked for memorability judgments during the memory test and found higher memorability judgments for LWF words relative to HWF items. Thus, the actual retrieval attempt provides the evidence needed to judge LWF words as more memorable.

The idea of a memorability-based rejection process bears similarity to research on the distinctiveness heuristic. Israel and Schacter (1997) demonstrated that false memory was significantly reduced when pictures were encoded along with the auditory presentation of the word as compared with when only auditory encoding occurred. This result presumably occurs because participants metacognitively use the absence of diagnostic, pictorial details in memory as evidence that if they had studied the item, they would have remembered it. LWF items are encountered less, which likely makes them more inherently distinctive than HWF items. Thus, a similar rejection process to the distinctiveness heuristic could be applied to LWF items more often than HWF items (Benjamin, 2003; Brown et al., 1977; Guttentag & Carroll, 1998).

Ghetti (2003) focused on a rejection process similar to the distinctiveness heuristic. In particular, she evaluated the ability to make the metacognitive judgment that if an item would have been studied, one would have remembered it (i.e., a "do not recall to reject" process; see Strack & Bless, 1994 for a similar idea). In Experiment 1, Ghetti manipulated item salience and asked participants to subjectively classify their rejection of items into one of three categories (one being the previously mentioned metacognitive process). A second mechanism was based on the feeling of a lack of familiarity. If an item lacked the familiarity to pass a certain criterion, the participant should report this process. Note that this option does not require an independent assessment of the subjective memorability of an item based on its salient characteristics (Ghetti, 2003). The third rejection process involved the recall of information from the encoding context that led to the rejection of a new item on a test list (e.g., Rotello & Heit, 2000). A recollection rejection mechanism, when used, has also been shown to reduce false memories (e.g., Brainerd, Reyna, & Estrada, 2006; Gallo, 2004; Schmid, Herholz, Brandt, & Buchner, 2010). Its use, however, appears to be dependent on numerous factors such as the use of related stimuli and test cues (Gallo, 2004; Light, 2012; Schmid et al., 2010) and explicit instructions to use the rejection process (e.g., Lampinen, Arnal, & Hicks, 2009; Rotello, Macmillan, & Van Tassel, 2000). In the current study, we took this approach of asking participants to classify the rejection

of LWF and HWF new items into one of these three categories (very similar to Ghetti, 2003, Experiment 1). With some exceptions (e.g., Ghetti, 2003; Kapucu, Macmillan, & Rotello, 2010; Lampinen, Meier, Arnal, & Leding, 2005; Lampinen, Odegard, & Neuschatz, 2004; Marsh et al., 2009; Migo, Montaldi, Norman, Quamme, & Mayes, 2009), there is relatively little research on subjective classifications of correct rejections when compared to the wealth of research using the Remember–Know procedure.

4. The current study

Despite the preliminary evidence associating LWF items with an enhanced perception of memorability, it is unknown whether participants rely primarily on a metacognitive, memorability assessment to correctly reject LWF items. Because LWF items do not have many pre-experimental associations, they could be rejected based on a subjective assessment that they simply lack the familiarity to have been studied. These lower levels of familiarity could be related to the decreased amount of implicit associative responses activated by LWF as compared to HWF items (Underwood & Freund, 1970). For example, *aardvark* may be rejected because it does not elicit feelings of familiarity due its lack of associations with other concepts. This process does not necessitate an independent metacognitive memorability assessment of *aardvark*. Previous researchers have only focused on the increased memorability for LWF items and thus did not directly assess other potential rejection processes (Benjamin, 2003; Guttentag & Carroll, 1998). Our primary goal in Experiment 1 was to assess whether the increased correct rejection rate for LWF items is related to a subjective judgment of increased memorability or to a lack of familiarity. We also included a recall-to-reject option in Experiment 1 as it has traditionally been examined in studies on rejection processes (e.g., Ghetti, 2003). We did not expect to find WF differences due to the lack of a basis to use such a process (e.g., rearranged word pairs; Light, 2012). We also sought to determine whether the rejection processes associated with the increased rejection of LWF items are confounded with the rejection processes for a word characteristic that is highly correlated with WF (i.e., context variability; Experiment 2). Context variability represents the number of contexts in which an item occurs (Steyvers & Malmberg, 2003). Even though the two variables are highly correlated, items of a particular WF may also differ in their context variability. In Experiment 3, we examined whether the differences in rejection strategies would still apply in pure lists (i.e., WF manipulated between-subjects). Overall, we sought to systematically investigate the subjective decision processes that contribute to the correct rejection portion of the WF mirror effect (and the report of such processes in general). To our knowledge, these experiments represent the first extensive experimental investigation into this topic.

5. Experiment 1

5.1. Method

5.1.1. Participants

Twenty-five undergraduate students from the University of Georgia participated in exchange for course credit. The participants were tested individually in a session that lasted approximately 30 min.

5.1.2. Materials and procedure

We selected a total of 120 words from the Kuĉera and Francis (1967) word norms. All of the items were nouns. Sixty were LWF items (M = 10 per million; range 9–11 per million) and 60 were HWF (M = 88.1 per million; range from 80 to 100 per million). These items were controlled on other relevant word characteristics (length, number of syllables, concreteness, and imagery) based on values from Kuĉera and Francis. Participants were instructed that they would see words appear sequentially on the computer screen and that they should study the words for a future, unspecified memory test. Afterwards, the experimenter reiterated the instructions to participants before initiating the study phase. The computer software randomly selected 60 total words from the pool of 120 to be studied. Thirty of these words were LWF and 30 were HWF. These words were presented for 2 s each.

As suggested by others (e.g., Migo et al., 2012), we have included our test instructions (please see Appendix A). The reader should note that additional *verbal* examples were given to participants for each subjective classification beyond what was in the *written* instructions. In summary, the participants were told that they would first decide whether an item was studied or not studied by pressing either the "Old" or "New" key on the keyboard. If they claimed that the word was studied, they would decide *why* they claimed the item was old. If they could recall some specific detail of when they saw the item, they were to indicate this by pressing a key to indicate they "Remembered" the item. If they recognized the word because it felt familiar (but with no specific details of the experimental context), they were instructed to press the "Know" key. If they pressed the old key, but were simply guessing (with no recollection or familiarity), they were instructed to press the "Guess" key. The "Guess" response was included mainly to equate these response options with those for the new items (i.e., three for each; Hicks & Marsh, 1999). Otherwise, these instructions were similar to those used in past experiments (e.g., Marsh et al., 2009).

If participants claimed they had *not* studied the item, they were also to indicate *why* they made this decision. If the item simply did not feel familiar, they should press the key indicating "Lacks Familiarity". If they recalled some detail from the

Table 1
Experiment 1 (WF manipulated within-subjects): Overall hit and correct rejection rates and the subjective classifications of the hits and correct rejections for LWF and HWF items.

Measure	LWF *M* (*SE*)	HWF *M* (*SE*)
Hit rate	.731 (.025)	.692 (.029)
Remember	.447 (.046)	.397 (.036)
Know	.201 (.030)	.175 (.021)
Guess	.083 (.014)	.120 (.018)
Correct rejection rate	.795 (.021)	.681 (.029)
Would have remembered	.212 (.029)	.093 (.017)
Lacks familiarity	.444 (.035)	.483 (.032)
Recall to reject	.139 (.033)	.105 (.023)
Conditionalized would have remembered	.267 (.036)	.137 (.023)
Conditionalized lacks familiarity	.563 (.044)	.714 (.040)
Conditionalized recall to reject	.170 (.041)	.149 (.029)

study context that led them to reject the test item, they were to indicate that by pressing the "Recall to Reject" key. Finally, if they believed that the item was so memorable/distinctive that it would have been remembered, they were to press a key indicating that they "Would Have Remembered" it had they studied it. Participants were given written and verbal examples of each type of rejection classification. The researcher verbally reiterated all of the instructions and made sure that the participant understood all of the decisions and classifications to be made. Once the experimenter was confident that the participant understood the instructions, the testing session began. The computer software randomly presented all 120 words. Of the 60 distractor items, half were LWF and half were HWF.

5.2. Results

5.2.1. Hit rates

A conventional alpha level of 5% was adopted for all statistical tests across all experiments. Refer to Table 1 for the relevant means and standard errors from Experiment 1. The overall hit rate was higher for LWF items than HWF items, $t(24) = 2.357$, $p = .027$, $d = .472$. We analyzed the raw proportions of hits labeled Remember, Know, and Guess in a 2 (WF) $\times$ 3 (Response Type) repeated measures ANOVA. Importantly, the interaction was significant, $F(2, 48) = 4.208, p = .021, \eta_p^2 = .149$. To explore the nature of this interaction, we compared the proportion of responses for each response option for HWF and LWF items. The proportion of Remember responses was higher for LWF items than HWF items, $t(24) = 2.615$, $p = .015$, $d = .526$.[1] The proportion of Guess responses, however, was higher for HWF items than LWF items, $t(24) = 2.477$, $p = .021$, $d = .497$. The proportion of Know responses did not differ across high and low WF items, $t(24) = 1.234$, $p = .229$.[2]

5.2.2. Correct rejection rates

While most research on the WF mirror effect has focused on the false alarm data, the correct rejection rate is simply the complement of the false alarm rate and thus the results are statistically equivalent. We chose to analyze correct rejections because we were more interested in why people correctly identify items as new.[3] The correct rejection rate was higher for LWF items than HWF items, $t(24) = 4.959$, $p < .001$, $d = .995$. Next, we analyzed the raw proportion of response types to correct rejections. In a 2 (WF) $\times$ 3 (Response Type) repeated measures ANOVA, the interaction was significant,

[1] We also analyzed Remember, Know, and Guess (Experiment 1 only) responses conditionalized only on hits. We reported the unconditionalized measures as this is common practice in the previous studies related to Remember–Know responses for low and high WF items (e.g., Cook et al., 2006; Joordens & Hockley, 2000). We believe that this measure better captures the amount of recollection and familiarity for *all* studied items, and not just those that were identified as having been studied. Using a conditionalized measure would result in unequal denominators across participants. For example, a participant that did not recognize many items could nevertheless have a high rate of conditionalized recollection. Thus, we believe the unconditionalized measure is the most appropriate way to partition Remember and Know responses. We do, however, analyze both the conditionalized and the unconditionalized measures for the rejection reports as it is more novel and the focus of the current paper. As reported in the results, these two types of analyses were largely consistent.

[2] We analyzed familiarity for studied items under an independence assumption (dividing the amount of Know claims by 1 minus the amount of Remember claims; see Jacoby, 1998). This manner of analysis is often used when it is believed that recollection and familiarity estimates are not independent and thus familiarity estimates are biased. There were no significant effects in this measure across all three experiments, suggesting that, at least in this manner of analysis, there was no difference in familiarity for studied low and high WF items. As this was not the focus of the study, these data are not reported as to keep the discussion more focused on rejection processes.

[3] We have focused on the subjective classifications associated with correct recognition decisions. Others, however, have reported these data for incorrect decisions (e.g., Remember–Know classifications to false alarms; e.g., Cook et al., 2006). While potentially informative, we do not report Remember-Know responses to false alarms as this is not the focus of the paper. While our focus is on rejection classifications, we did not report the analysis for rejection reports associated with misses primarily to keep the current discussion focused and because it is not clear what these data represent. We did include these means and standard errors in Appendix B (Experiment 1) and Appendix C (Experiments 2 and 3). While it may be tempting to make conclusions from this data, the pattern of results was not completely consistent across experiments. Also, the percentage of Would Have Remembered reports to misses was low and should be interpreted with caution. The detailed analysis is available upon request.

$F(2, 48) = 6.138, p = .004, \eta_p^2 = .204$. The proportion of Would Have Remembered responses was higher for LWF items than HWF items, $t(24) = 4.381$, $p < .001$, $d = .879$. The proportion of Lacks Familiarity responses, $t(24) = 1.210$, $p = .238$, and Recall to Reject responses, $t(24) = 1.588$, $p = .125$, did not differ across LWF and HWF items. We also conditionalized response types on just those items correctly rejected. For this measure, there were more conditionalized Would Have Remembered responses for LWF than HWF items, $t(24) = 3.889$, $p = .001$, $d = .778$. There were, however, more conditionalized Lacks Familiarity responses for HWF items than LWF items, $t(24) = 4.186$, $p < .001$, $d = .836$. The conditionalized Recall to Reject responses did not differ across LWF and HWF items, $t(24) = .758$, $p = .456$.

5.3. Discussion

We replicated the classic WF mirror effect in the hit and correct rejection rates and the finding that there are more Remember responses for LWF items than HWF items (e.g., Joordens & Hockley, 2000). With regards to reported rejection processes, LWF items were given more Would Have Remembered responses than HWF items, presumably due to their increased sense of memorability. In contrast, HWF items were given more conditionalized Lacks Familiarity responses than LWF items. Thus, when HWF items were correctly rejected, the decision was more likely to be related to an assessment of familiarity and not a perceived memorability process. Recollection rejection levels were low overall with no difference between the two classes of WF. This finding is consistent with the idea that there have to be very specific circumstances present in order to elicit differences in recollection rejection (e.g., Brainerd et al., 2006; Gallo, 2004; Light, 2012; Schmid et al., 2010). The results from Experiment 1 suggest that the WF mirror effect is not solely driven by the decision processes acting on memory for old items that are judged in the recognition test. An additional, important component of the WF mirror effect is found in participants' subjective classification associated with the rejection of new items that also differ in their normative WF.

6. Experiment 2

In Experiment 2, we sought to elucidate whether similar effects would be obtained for items varying in WF when controlling for context variability (CV). The measure of WF and CV created by Steyvers and Malmberg (2003) originates from the Touchstone Applied Science Associates (TASA) corpus (see Landauer, Foltz, & Laham, 1998 for complete details). Previous studies examining WF when controlling for CV have shown that these variables produce independent mirror effects (Cook et al., 2006; Steyvers & Malmberg, 2003). In addition, both low WF and low CV words are independently given more Remember responses than high WF and high CV words (Cook et al., 2006). While this is a different measure of WF than used in the first experiment, conceptually these measures are similar. In Experiment 2, we crossed WF with CV to create four different classes of stimuli. Thus, we could examine the different rejection mechanisms for these two constructs. It is unknown whether the differential memorability of WF items is confounded with CV and vice versa. It is conceivable that the perception of increased subjective memorability of LWF items is due to the fact they are not encountered in a wide variety of contexts. In this experiment (and the next), we removed the Guess and Recall-to-Reject options due to their low usage and lack of differences in Experiment 1.

6.1. Method

6.1.1. Participants

Thirty-five undergraduate students from the University of Georgia participated in exchange for course credit. The participants were tested individually in a session that lasted approximately 30 min. We increased the sample size from Experiment 1 due to the smaller number of items per class. One participant was excluded due to a software error.

6.1.2. Materials and procedure

For this experiment, we selected items from the Steyvers and Malmberg (2003) corpus. This set of words was grouped into four classes of items: LWF/low context-variability (LCV), LWF/high-context variability (HCV), HWF/LCV, HWF/HCV. We removed six items from each of the four categories due to their taboo nature. This left us with a total of 268 words (67 per class). The computer software randomly chose 20 words from each class to present during the study phase for a total of 80 words presented. Otherwise, the study phase was the same as in the first experiment. The test instructions were identical to Experiment 1 with the exception that there was no Guess or Recall-to-Reject option. At test, the computer software presented all items from the study phase as well as 80 new items (20 from each class) randomly chosen from the remaining stimuli.

6.2. Results

6.2.1. Hit rates

Please refer to Table 2 for all means and standard errors from Experiment 2. We do not report the interaction effects in Experiment 2 and 3 because there were no statistically significant interactions. One interaction was marginally significant in Experiment 3 and is thus reported. In a 2 (WF) $\times$ 2 (CV) repeated measures ANOVA on hits, we found main effects of both

Table 2
Experiment 2 (CV and WF manipulated within-subjects): Overall hit and correct rejection rates and the subjective classifications of the hits and correct rejections for each of the four classes of WF and CV items and averaged separately over each of the two classes of LWF, HWF, LCV, and HCV items.

Measure	LWF/LCV *M* (*SE*)	LWF/HCV *M* (*SE*)	HWF/LCV *M* (*SE*)	HWF/HCV *M* (*SE*)	Avg LWF *M* (*SE*)	Avg HWF *M* (*SE*)	Avg LCV *M* (*SE*)	Avg HCV *M* (*SE*)
Hit rate	.734	.632	.638	.582	.683	.610	.686	.607
	(.024)	(.024)	(.023)	(.019)	(.020)	(.017)	(.019)	(.018)
Remember	.510	.363	.382	.285	.437	.334	.446	.324
	(.029)	(.028)	(.028)	(.024)	(.026)	(.022)	(.025)	(.022)
Know	.224	.269	.256	.297	.246	.277	.240	.293
	(.026)	(.025)	(.023)	(.021)	(.021)	(.020)	(.023)	(.020)
Correct rejection rate	.888	.829	.757	.742	.859	.750	.823	.786
	(.014)	(.022)	(.029)	(.031)	(.015)	(.027)	(.018)	(.023)
Would have remembered	.388	.265	.271	.172	.327	.221	.329	.218
	(.030)	(.031)	(.030)	(.028)	(.027)	(.026)	(.024)	(.025)
Lacks familiarity	.500	.565	.487	.571	.532	.529	.493	.568
	(.028)	(.034)	(.032)	(.036)	(.027)	(.029)	(.026)	(.030)
Conditionalized would have remembered	.434	.322	.351	.228	.378	.290	.393	.275
	(.032)	(.037)	(.035)	(.036)	(.031)	(.032)	(.028)	(.032)

WF, $F(1,33) = 13.460, p = .001, \eta_p^2 = .290$, and CV, $F(1,33) = 19.763, p < .001, \eta_p^2 = .375$. These main effects demonstrate higher hit rates for both LWF and LCV items. We analyzed raw Remember and Know responses in two separate 2 (WF) × 2 (CV) repeated measures ANOVAs. For Remember responses, we found main effects of both WF, $F(1,33) = 22.360, p < .001, \eta_p^2 = .403$, and CV, $F(1,33) = 39.710, p < .001, \eta_p^2 = .546$. These results indicate more Remember responses for LWF and LCV items than HWF and HCV items. For Know responses, we also obtained main effects for WF, $F(1,33) = 4.146, p = .050, \eta_p^2 = .112$, and CV, $F(1,33) = 4.878, p = .034, \eta_p^2 = .129$, which indicate more Know responses for the HWF and HCV items than LWF and LCV items.

6.2.2. Correct rejection rates

In a similar 2 (WF) × 2 (CV) repeated measures ANOVA on correct rejections, we found a main effect of WF, $F(1,33) = 22.527, p < .001, \eta_p^2 = .406$, and a marginal main effect of CV, $F(1, 33) = 3.873$, $p = .058$. The main effect of WF and the marginal effect of CV indicate higher correct rejection rates for LWF and LCV. We then analyzed each raw proportion of rejection responses in two separate ANOVAs. For Would Have Remembered responses, we found main effects of both WF, $F(1,33) = 15.808, p < .001, \eta_p^2 = .324$, and CV, $F(1,33) = 57.276, p < .001, \eta_p^2 = .634$. This result demonstrates more Would Have Remembered responses for LWF and LCV items than HWF and HCV items. For Lacks Familiarity responses, the main effect of CV was significant, $F(1,33) = 9.059, p = .005, \eta_p^2 = .205$, but the main effect of WF was not, $F(1, 33) = .022$, $p = .833$. There were more Lacks Familiarity responses for HCV items than LCV items. In terms of conditionalized Would Have Remembered responses, there were main effects of both WF, $F(1,33) = 10.138, p = .003, \eta_p^2 = .235$, and CV, $F(1,33) = 34.196, p < .001, \eta_p^2 = .509$.[4] The two main effects demonstrate higher Would Have Remembered rates for LWF and LCV items.

6.3. Discussion

We replicated Cook et al. (2006) and Steyvers and Malmberg (2003) in finding mirror effects for both WF and CV separately (marginally in the case of rejection rates for CV). The hit rate for both effects was related to increased Remember responses (Cook et al., 2006). The amount of Know responses was higher for the HWF and HCV items than the LWF and LCV items. In terms of subjective rejection processes, we found more raw and conditionalized Would Have Remembered responses for both LWF and LCV items. Thus, the increased perception of memorability for WF is not confounded with CV and vice versa. For raw Lacks Familiarity responses, only the effect of CV was present such that HCV items were given more Lacks Familiarity responses than LCV items. These results replicate Experiment 1 in that the correct rejection of LWF items is related to the perception of their memorability. In general, we found CV and WF produced similar, but independent patterns of recognition decisions, including subjective classifications.

7. Experiment 3

In Experiment 3, we wanted to know whether the differences in rejection classifications would still occur when we manipulated WF between-subjects. The WF mirror effect is often found in both mixed and pure lists (e.g., Clark & Burchett,

[4] We did not analyze a conditionalized measure of Lacks Familiarity responses in Experiments 2 and 3 because this measure is linearly dependent with the conditionalized measure of Would Have Remembered responses. Because there were only two response options in these experiments, all correct rejections must be labeled as either Lacks Familiarity or Would Have Remembered. Thus, the sum of the two measures would be one and the ANOVA results would be the exact same.

1994; Dorfman & Glanzer, 1988; Gorman, 1961; McCormack & Swenson, 1972) and the hit rate is due to recollection in both cases (Dewhurst et al., 2004). Dewhurst et al., however, did not find the correct rejection component of the WF mirror effect in a pure list manipulation. Malmberg and Murnane (2002) also found that the false alarm rate for LWF decreased with the presence of more studied HWF items. It is possible, thus, that participants are not sensitive to memorability differences in new items when presented with all LWF or HWF items because relative differences between the memorability for these items is not diagnostic of list membership. If this is the case, there should be no differences in Would Have Remembered responses between LWF and HWF words. As in Experiment 2, CV was still manipulated within-subjects. We chose a between-subjects design to prevent exposure to any items of contrasting WF during the experiment.

7.1. Method

7.1.1. Participants

Seventy undergraduate students from the University of Georgia participated in exchange for course credit. Thirty-five participants were randomly assigned to a LWF condition and the other 35 were assigned to a HWF condition. One participant was excluded due to a software error. The participants were tested individually in a session that lasted approximately 30 min.

7.1.2. Materials and procedure

As in Experiment 2, we selected items from the Steyvers and Malmberg (2003) corpus. For this experiment, however, we only presented LWF words at study and test in the first condition while only HWF words were presented in the second condition. CV was manipulated within-subjects such that all participants encountered LCV and HCV items at study and test. Thus, in the first condition, participants studied 30 LWF/LCV items and 30 LWF/HCV items and then were tested on these 60 words and 30 new words from each of these two classes of items (120 total test items). In the second condition, participants studied 30 HWF/LCV items and 30 HWF/HCV items and then were tested on these 60 words and 30 new words from each class. Otherwise, the study and test procedure were the same as in Experiment 2.

7.2. Results

7.2.1. Hit rates

Please refer to Table 3 for all means and standard errors from Experiment 3. Analyses were conducted similarly to Experiment 2 with the exception that WF was compared between-subjects. In a 2 (WF) × 2 (CV) mixed ANOVA on hits, the main effect of CV was significant, $F(1,67) = 31.972, p < .001, \eta_p^2 = .323$, but the main effect of WF was not, $F(1, 67) = 1.173$, $p = .283$. This main effect demonstrates higher hit rates for LCV items but *not* for LWF items. The pattern of the WF hit rate data, however, was numerically consistent with the typical WF mirror effect. We analyzed the raw proportions of Remember and Know responses in two separate 2 (WF) × 2 (CV) mixed ANOVAs. For Remember responses, we found a main effect of CV, $F(1,67) = 42.340, p < .001, \eta_p^2 = .387$, but not WF, $F(1, 67) = .140$, $p = .709$. This main effect demonstrates higher raw Remember rates for LCV items but *not* for LWF items relative to their HCV and HWF counterparts, respectively. For Know responses, we also obtained a main effect of CV, $F(1,67) = 8.164, p = .006, \eta_p^2 = .109$, but no main effect of WF, $F(1, 67) = .283$, $p = .596$. This main effect demonstrates more Know responses for the HCV items than the LCV items.

Table 3
Experiment 3 (WF manipulated between-subjects and CV manipulated within-subjects): Overall hit and correct rejection rates and the subjective classifications of the hits and correct rejections for each of the four classes of WF and CV items and averaged separately over each of the two classes of LWF, HWF, LCV, and HCV items.

Measure	LWF/LCV *M* (*SE*)	LWF/HCV *M* (*SE*)	HWF/LCV *M* (*SE*)	HWF/HCV *M* (*SE*)	Avg LWF *M* (*SE*)	Avg HWF *M* (*SE*)	Avg LCV *M* (*SE*)	Avg HCV *M* (*SE*)
Hit rate	.714 (.022)	.631 (.028)	.676 (.022)	.600 (.026)	.673 (.023)	.638 (.022)	.695 (.016)	.615 (.019)
Remember	.435 (.035)	.302 (.031)	.413 (.028)	.294 (.030)	.369 (.030)	.354 (.026)	.424 (.022)	.298 (.021)
Know	.278 (.029)	.329 (.032)	.256 (.023)	.297 (.021)	.304 (.028)	.284 (.025)	.271 (.019)	.317 (.021)
Correct rejection rate	.860 (.025)	.780 (.025)	.780 (.020)	.744 (.022)	.820 (.021)	.762 (.020)	.819 (.015)	.762 (.017)
Would have remembered	.357 (.031)	.205 (.028)	.322 (.021)	.182 (.026)	.281 (.027)	.252 (.025)	.339 (.021)	.193 (.019)
Lacks familiarity	.503 (.035)	.576 (.037)	.458 (.031)	.562 (.033)	.539 (.034)	.510 (.030)	.480 (.023)	.569 (.025)
Conditionalized would have remembered	.419 (.036)	.270 (.038)	.413 (.038)	.249 (.026)	.345 (.035)	.331 (.034)	.416 (.026)	.259 (.026)

7.2.2. Correct rejection rates

In terms of correct rejections, there was a main effect of CV, $F(1,67) = 24.180, p < .001, \eta_p^2 = .265$, and WF, $F(1,67) = 4.066, p = .048, \eta_p^2 = .057$. These effects demonstrate higher rejection rates for LWF and LCV items. The interaction approached significance, $F(1, 67) = 3.380$, $p = .070$. This marginal interaction should be interpreted cautiously. The correct rejection difference between LWF and HWF items was numerically equivalent to the correct rejection difference between LCV and HCV items. The marginal interaction is likely a result of the fact that the within-subjects variable (CV) has a lower standard error than WF, which was manipulated between-subjects. We next analyzed each raw proportion of rejection responses in two separate ANOVAs. For Would Have Remembered responses, we found a main effect of CV, $F(1,67) = 85.075, p < .001, \eta_p^2 = .559$, but no main effect of WF, $F(1, 67) = .611$, $p = .437$. This result demonstrates higher raw Would Have Remembered rates for LCV items but *not* for LWF items. For Lacks Familiarity responses, we also obtained a main effect for CV, $F(1,67) = 25.661, p < .001, \eta_p^2 = .277$, but no main effect of WF, $F(1, 67) = .422$, $p = .518$. This main effect demonstrates more Lacks Familiarity responses for the HCV items than LCV items. In terms of conditionalized Would Have Remembered responses, there was a main effect of CV, $F(1,67) = 66.372, p < .001, \eta_p^2 = .498$, but no main effect of WF, $F(1, 67) = .081$, $p = .776$. This main effect demonstrates higher Would Have Remembered rates only for LCV items.

7.3. Discussion

We replicated the finding from Experiment 2 and previous studies (Cook et al., 2006; Steyvers & Malmberg, 2003) that demonstrates a mirror effect such that LCV items have higher hit rates and lower false alarm rates than HCV items. Here, the correct rejection portion of the CV mirror effect was significant and not marginally significant. In addition, there were more Remember and Would Have Remembered responses and less Know and Lacks Familiarity responses for LCV items. The results further support the notion that the CV mirror effect is related to increased recollection for the studied items and an increased use of a memorability-based judgment for the new items. Contrasting with Experiments 1 and 2, we failed to find a statistical advantage in the hit rates for LWF items as compared with HWF items. In addition, there were no differences in Remember and Know responses between these two classes of items (the numerical difference between Remember responses for LWF and HWF items was nearly nonexistent). We did find, however, that correct rejection rates were higher for LWF items. Despite this advantage, we did not find any differences in Would Have Remembered or Lacks Familiarity responses between LWF and HWF items (there was only a very slight numerical difference). Thus, even though we did replicate the correct rejection portion of the WF mirror effect, participants no longer classified LWF items as being words they Would Have Remembered more often than HWF items. This finding suggests that the relative/contextual distinctiveness across items influences the report of a memorability-based assessment associated with the rejection of unstudied items. As will be explained more in Section 8, the failure to find some of these effects may be partially due to the reduced power of using a between-subjects as compared with a within-subjects design (though we will argue this is not the primary determinant). We had .54 power to detect medium-sized effects and .90 power to detect a large sized-effects (Faul, Erdfelder, Buchner, & Lang, 2009).

8. General discussion

The primary goal of the current study was to provide a more extensive examination of the reported mechanisms responsible for the WF mirror effect currently found in the literature. Researchers have traditionally framed mirror effects as a tradeoff between hit rates and false alarm rates (e.g., Glanzer et al., 1993). A particularly compelling suggestion derived from the current study is that participants do not only reduce false alarm rates by mechanisms that support making old decisions. Rather, participants capitalize on additional processes to classify lures in a recognition memory test as new. These processes are related to general memory mechanisms that support successful recognition (i.e., familiarity and perceived memorability). However, and as will be discussed subsequently, these processes are subjectively reported and may not always be directly and causally linked to the actual recognition decision.

8.1. Examination of hit rate findings

We replicated the hit rate and recollection advantage for LWF and LCV items when these variables were manipulated within-subjects (participants encountered both low and high WF and CV items; e.g., Reder et al., 2000). We did not find statistical evidence for the LWF hit rate advantage in pure-list manipulations, which is inconsistent with prior research (e.g., Dorfman & Glanzer, 1988; Gorman, 1961; McCormack & Swenson, 1972). A closer inspection of the data shows that the numerical hit rate advantage was present and suggests that the absence of a significant effect may be due to a lack of power in our between-subjects design. In the first two experiments, the hit rate advantage for LWF items was numerically smaller than the correct rejection advantage for LWF items. One should note, however, that when the same stimuli were manipulated in a within-subjects fashion (Experiment 2), the hit rate difference was nearly twice as large as when manipulated in a between-subjects manner (Experiment 3). Past pure-list studies of the WF mirror effect have often used a blocked, within-subjects design (e.g., Clark & Burchett, 1994; Gorman, 1961; McCormack & Swenson, 1972; but see Dewhurst et al., 2004). This exposure to both LWF and HWF items could have affected responding (McCabe & Balota, 2007). If one takes the position,

however, that the hit rate advantage exists, but we did not have the power to find it, then the elimination of the differences in Remember responses may seem puzzling. Perhaps participants developed an expectation of how much recollection to expect based on the study list. This *expectancy heuristic* (McCabe & Balota, 2007) may have resulted in the participants altering their judgment of what constitutes a Remember response, and thus they assigned fewer Remember responses when only LWF items were studied.

8.2. Summary of correct rejection findings

Our primary variable of interest in this study was the correct rejection portion of the mirror effect. In within-subjects (mixed-list) manipulations, we found mirror effects, but we also found that correct rejections to LWF items were more likely to be attributed to a memorability process and not to a lack of perceived familiarity or a recollection rejection mechanism. This finding is consistent with research that has demonstrated that at test, people are aware that LWF items are more memorable, and thus, a lack of memory for those items is indicative of not having studied them (Benjamin, 2003; Brown et al., 1977; Guttentag & Carroll, 1998). We, however, directly compared this memorability process with other types of rejection processes using subjective classifications. Participants could have conceivably rejected LWF items based on an assessment that they do not have high levels of familiarity, and thus do not pass a criterion for having been studied (Arndt & Reder, 2002; Guttentag & Carroll, 1994; Joordens & Hockley, 2000; Reder et al., 2000 Rugg et al., 1995). Our data suggest that even when afforded the opportunity to respond with a Lacks Familiarity option, the rejection advantage for LWF items is related to participants' increased, metacognitive perception of memorability. We also found that the correct rejection advantage for LWF items was not confounded with CV (Cook et al., 2006). This independence was also observed for subjective classifications in that people were still more likely to give LWF items a Would Have Remembered response even when CV was controlled. Thus, there is something inherently more memorable about LWF items that is not attributable to the variety of contexts in which they occur.

We also replicated the correct rejection advantage for LWF items in pure lists (e.g., Clark & Burchett, 1994; Dorfman & Glanzer, 1988; Gorman, 1961; McCormack & Swenson, 1972; but see Dewhurst et al., 2004). This result suggests that there need not be contrasting HWF items presented at study or test to find the correct rejection advantage (but see Malmberg & Murnane, 2002) and is consistent with the notion that stimulus-based mirror effects (including WF) are not subject to criterion changes in the same manner as strength-based mirror effects (Stretch & Wixted, 1998). The pattern of subjective classifications, however, did change in the pure-list manipulation. Participants no longer assigned LWF items more Would Have Remembered responses as compared with HWF items. While this result may be surprising given the fact that the correct rejection effect remained, it is not inconsistent with past research on WF memorability strategies because they only involved mixed-list methods (Benjamin, 2003; Brown et al., 1977; Guttentag & Carroll, 1998). The fact that the differences in subjective classifications remained for CV (which was manipulated within-subjects) is further evidence that the lack of a WF effect was due to the between-subjects manipulation. In addition, a lack of power is likely not the primary explanation for this null effect. The numerical difference between Would Have Remembered responses given to LWF and HWF items was much larger in Experiment 2 than in Experiment 3 (the same stimuli were used in both experiments). At the very least, the between-subjects manipulation greatly decreased (if not eliminated) perceived memorability differences between unstudied LWF and HWF items.

8.3. Closer examination of the correct rejection findings

The infrequent occurrence of LWF items makes them inherently more distinctive and more memorable than HWF words. This inherent or secondary distinctiveness (Schmidt, 1991) likely accounts for the increased ability to reject these stimuli. When only LWF items are studied, they maintain their inherent distinctiveness, but lose their contextual distinctiveness due to the lack of contrasting HWF items. Perhaps the inherent distinctiveness is enough for people to reject LWF items in pure lists. The contextual distinctiveness, however, may change participants' definition of what constitutes a subjective Would Have Remembered classification. Thus, similar to Remember responses, participants *expect* items to be more memorable (i.e., higher average memorability) when only encountering LWF stimuli (see McCabe & Balota, 2007 for a similar idea). When both LWF and HWF items are present, the *expected* average memorability would not be as high, thus resulting in more Would Have Remembered judgments for LWF items. This result suggests that a Would Have Remembered judgment is not just based on the innate characteristics of the item itself, but can also be influenced by the specific test context. This explanation is similar to the functional account of recollection (Bodner & Lindsay, 2003; Gruppuso, Lindsay, & Kelley, 1997). This account states that the judgment of whether recollection is present depends on the context and when Remember judgments would be *criterial* in making a recognition decision. The same argument could be made for rejection processes in that context could affect when a Would Have Remembered would be criterial for a rejection decision (e.g., the presence or absence of HWF items). Whether this manipulation only changed the criterion for what constitutes a Would Have Remembered response or whether it more fundamentally changed the type of information considered by the participants is not completely evident. Future research should continue to clarify these effects.

8.4. The general use of subjective classification of memory processes

The debate about the use of subjective Remember–Know judgments to make conclusions about underlying recollection and familiarity processes could apply to the use of such classifications to study rejection processes. For example, it is possible that participants may have interpreted Know and Lacks Familiarity responses as representing less confident (i.e., "weaker") decisions as compared to Remember and Would Have Remembered responses (see Donaldson, 1996; Dunn, 2004; and Wais et al., 2008 for a similar argument related to Remember–Know responses). In addition, Experiment 3 demonstrated that the correct rejection portion of the mirror effect remained, whereas the statistical differences for Would Have Remembered responses given to LWF versus HWF items were eliminated. Despite these results, participants judged LWF items as more memorable in the context of HWF items, even when accounting for CV. As with the Remember–Know paradigm (e.g., Migo et al., 2012), we believe that if used properly, these subjective classifications can be effectively utilized. Future research on these rejection classifications should parallel those that have been done with the Remember–Know paradigm including accounting for confidence levels (e.g., McCabe, Geraci, Boman, Sensenig, & Rhodes, 2011; Williams et al., 2013), using different procedures (e.g., think aloud protocols; Lampinen et al., 2005; McCabe et al., 2011), and examining the neural correlates associated with these rejection processes (e.g., Curran, 2004; Skinner & Fernandes, 2007; Spaniol et al., 2009).

8.5. Limitations, future directions, and conclusions

Because the study of rejection processes is somewhat exploratory, more clarification is needed. For example, there may be other processes aside from the three on which we focused that are important not only in mirror effects, but recognition memory in general. Within the processes we studied, more differentiation is needed (i.e., the exact difference between a Would Have Remembered and Lacks Familiarity classification). We believed that requiring participants to make mutually-exclusive rejection responses would force them to choose the strategy that best represents the information used when making the judgment. We do acknowledge, however, that our methodology of requiring participants to select only one rejection response is a potential limitation. Related to studied material, items that are claimed to be "Remembered" may also contain "familiarity" (e.g., Jacoby, 1998). In a similar manner, items could be rejected based on multiple processes. Using mutually-exclusive rejection responses limits our knowledge about the complete distribution of memorial characteristics examined when rejecting nonstudied items. In addition, we have assumed that the participants equally weighed the rejection response options. It is possible, however, that the differential use of these rejection processes could be based on an initial metacognitive assessment of the memorability for the item. If an item is perceived as memorable, but there is no memory for the item, the metacognitive option could be chosen. If, however, the item does not feel familiar but lacks an assessment of high memorability, a lack of familiarity option might be chosen. While a lack of familiarity might also contribute to the increased rejection of LWF items, the metacognitive judgment might be the more diagnostic response and thus would be reported. In a related sense, Would Have Remembered judgments could simply reflect more confident rejection decisions than those associated with a Lacks Familiarity report. Think-aloud protocols/self-reports (Lampinen et al., 2005; McCabe et al., 2011) and the assessment of another's subjective classifications (Williams et al., 2013) could be used to clarify the use and reporting of these rejection processes.

In conclusion, our study was not only the first extensive examination of the subjective reports of rejection processes in the WF mirror effect, but one of the few extensive investigations of rejection processes in general. Our results suggest that participants do report using a perceived memorability process more to reject LWF as compared to HWF items (independent of the lexical characteristic context variability), but they are only consciously aware of this when studied in the context of HWF items. Future work should continue to clarify the use of rejection processes not only in word frequency mirror effects, but also other manipulations that may lead to differences in the use of these processes.

Appendix A

A.1. Test instructions for Experiment 1

You have just studied the list of words. In this next and final phase of the experiment, your memory for the words will be tested. Some of the words on this test will be words that you studied and some them will be words that were not studied (they are brand new). For each word during this memory test, you will be deciding whether you believe it was either studied or is brand new. Please press the OLD key with your RIGHT INDEX finger if you believe test word was studied and press the NEW key with your LEFT INDEX finger if you believe the test word was not studied earlier.

Following your OLD or NEW response, you will be asked to make another response for each word. The instructions that follow are important for understanding how to respond, so please read them carefully.

If you called the word OLD, you will asked to make another judgment between 3 choices. When people like yourself complete a memory test like the one you are about to complete, they typically report remembering studying words based on 1 of 3 types of information, which are described below.

1. One remembers SPECIFIC DETAILS about the event or experience. One might recollect that the word came toward the beginning or end of the study list. One might recollect a particular reaction that accompanied the word. One might recollect what other words came right before or after the word. One might recollect that the word made you think of something else.
   For example, you may remember that you recently saw a movie because you can recollect specific details about the event like who you went with, or that you ate too much popcorn, or even your reaction to the movie itself. In other words, you remember what movie you saw because you can also RECOLLECT other details of the experience. A memory like this is referred to as REMEMBERING and is abbreviated "R" for remembering specific details.
2. One might not remember specific details, but nevertheless, the experience feels vaguely FAMILIAR. This does not mean that the event did not occur, only that we cannot place the event in its original context.
   For example, someone might respond quite confidently that he brushed his teeth 3 days ago, but he or she probably does not have any specific recollective details of the experience. A memory like this is referred to as KNOWING and is abbreviated "K" for knowing.
3. Sometimes one might have no recollective details for studying a word, nor have any familiarity for it, but nevertheless judge the word as OLD (studied). In this case, people might simply GUESS that a test word was old. This option is abbreviated G "Guess.

After considering the best way to characterize your memory for studying the word from the 3 types described above, you will press the corresponding key on the keyboard labeled K for "KNOW", R for "Remember Details", and G for "Guess" as shown below:

K R

G

You will also be asked to make one of three different judgments for NEW words. People typically claim that they have NOT studied particular word based on 1 of 3 types of information, which are described to you below.

1. The word did not seem or feel familiar because it lacked familiarity, thus one believes that it was not studied.
   This is similar to telling Person X that you do not remember meeting him/her because he or she just "Does not look familiar to you" and that you do not recognize him/her. This option is abbreviated LF for "Lacks Familiarity".
2. One may feel that they searched memory and recalled something about the study phase that lead them to believe the word was not studied. In this way, the test item can be rejected because one searches and may find similar items, but not the specific test item.
   For example, if the test word is TRAIN, you may search your memory of the test phase and remember that you studied PLANE. This leads you to believe that you did not study the similar word, TRAIN. This option is abbreviated RR because you can "Reject it if you Recall similar details from the study phase that lead you to reject it".
3. One thinks that the word is so distinctive that if it was studied, it would have been remembered, and because it is not remembered, it must not have been studied.
   For example, if your first name appeared on the test you might judge it as NEW because you feel that it would have been so distinctive to you that if you did study it, you would have remembered it. This option is abbreviated WHR for "Would Have Remembered it".

After considering the best way to characterize NOT having a memory for the word, you will press the corresponding key on the keyboard labeled LF for "Lacks Familiarity, RR for "Recall to Reject", and WHR for "Would Have Remembered It" as shown below:

LF RR

WHR

For clarification, if you think that a word was studied, you will press the OLD key. You will then decide whether you (1) KNOW it because it feels familiar to you, (2) REMEMBER specific details, or (3) GUESSED that it was old. Please respond by pressing the R, K, or G key, respectively.

If you think that a word was not studied, you will press the NEW key. You will then decide whether you judged it as new because it (1) LACKED FAMILIARITY to have been studied, (2) you can REJECT it because you RECALL detail from the study phase that caused you to reject the item or (3) you WOULD HAVE REMEMBERED STUDYING IT if it had been studied. Please respond by pressing the LF, RR, or WHR key, respectively.

Please ask the experimenter to clarify any questions that you have about the instructions that you just read.

Appendix B

Experiment 1 (WF manipulated within-subjects): Overall miss rates and the subjective classifications of the misses for LWF and HWF items.

Measure	LWF *M* (*SE*)	HWF *M* (*SE*)
Miss rate	.269 (.025)	.319 (.029)
Would have remembered	.047 (.001)	.047 (.001)
Lacks familiarity	.184 (.021)	.215 (.026)
Recall to reject	.039 (.011)	.047 (.023)
Conditionalized would have remembered	.169 (.030)	.150 (.029)
Conditionalized lacks familiarity	.699 (.046)	.696 (.051)
Conditionalized recall to reject	.132 (.033)	.154 (.038)

Appendix C

Experiment 2 (CV and WF manipulated within-subjects) and Experiment 3 (WF manipulated between-subjects and CV manipulated within-subjects): Overall miss rate and the subjective classifications of the misses for each of the four classes of WF and CV items and averaged separately over each of the two classes of LWF, HWF, LCV, and HCV items.

Measure	LWF/ LCV *M* (*SE*)	LWF/ HCV *M* (*SE*)	HWF/ LCV *M* (*SE*)	HWF/ HCV *M* (*SE*)	Avg LWF *M* (*SE*)	Avg HWF *M* (*SE*)	Avg LCV *M* (*SE*)	Avg HCV *M* (*SE*)
Experiment 2 miss rate	.266 (.024)	.368 (.024)	.362 (.023)	.418 (.019)	.317 (.020)	.390 (.017)	.314 (.019)	.393 (.018)
Would have remembered	.101 (.015)	.087 (.015)	.104 (.016)	.090 (.018)	.094 (.011)	.097 (.014)	.103 (.012)	.088 (.016)
Lacks familiarity	.165 (.018)	.281 (.025)	.257 (.025)	.328 (.023)	.223 (.018)	.293 (.021)	.211 (.018)	.304 (.022)
Conditionalized would have remembered	.373 (.045)	.248 (.039)	.301 (.046)	.215 (.041)	.310 (.033)	.258 (.037)	.337 (.036)	.231 (.038)
Experiment 3 miss rate	.286 (.022)	.369 (.028)	.324 (.022)	.400 (.026)	.327 (.023)	.362 (.022)	.305 (.016)	.385 (.019)
Would have remembered	.103 (.016)	.095 (.016)	.112 (.013)	.080 (.016)	.099 (.015)	.096 (.013)	.108 (.010)	.087 (.011)
Lacks familiarity	.183 (.020)	.274 (.024)	.211 (.021)	.320 (.025)	.228 (.021)	.266 (.022)	.198 (.015)	.297 (.018)
Conditionalized would have remembered	.361 (.050)	.261 (.040)	.351 (.042)	.201 (.039)	.311 (.041)	.276 (.038)	.356 (.032)	.231 (.028)

References

Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 830–842.

Benjamin, A. S. (2003). Predicting and postdicting the effect of word frequency on memory. *Memory & Cognition, 31*, 297–305.

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remembering" and a caution about purportedly nonparametric measures. *Memory & Cognition, 33*, 261–269.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563–580.

Brainerd, C. J., Reyna, V. F., & Estrada, S. (2006). Recollection rejection of false narrative statements. *Memory, 14*, 672–691.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency, and negative recognition. *Quarterly Journal of Experimental Psychology, 29*, 461–473.

Clark, S. E., & Burchett, R. E. (1994). Word frequency and list composition effects in associated recognition and recall. *Memory & Cognition, 22*, 55–62.

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.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42*, 1088–1106.

Dewhurst, S. A., Brandt, K. R., & Sharp, M. S. (2004). Intention to learn influences the word frequency effect in recall but not recognition memory. *Memory & Cognition, 32*, 1316–1325.

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 284–298.

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.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 120–128.
Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.
Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language, 48*, 722–739.
Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.
Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*, 546–567.
Gorman, A. N. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*, 23–29.
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*, 259–279.
Guttentag, R. E., & Carroll, D. (1994). Identifying the basis for the word-frequency effect in recognition memory. *Memory, 2*, 255–273.
Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory & Cognition, 26*, 951–958.
Hicks, J. L., & Marsh, R. L. (1999). Remember–know judgments can depend on how the question is asked. *Psychonomic Bulletin & Review, 6*, 117–122.
Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*, 577–581.
Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 3–26.
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.
Kapucu, A., Macmillan, N. A., & Rotello, C. M. (2010). Positive and negative remember judgments and ROCs in the plurals paradigm: Evidence for alternative decision strategies. *Memory & Cognition, 38*, 541–554.
Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press.
Lampinen, J., Arnal, J. D., & Hicks, J. L. (2009). Prospective person memory. In M. R. Kelley (Ed.), *Applied memory* (pp. 167–184). Hauppauge, NY, US: Nova Science Publishers.
Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 954–963.
Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1997). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology, 16*, 181–224.
Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 332–342.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.
Light, L. L. (2012). Dual-process theories of memory in old age: An update. In M. Naveh-Benjamin & N. Ohta (Eds.), *Memory and aging: Current issues and future directions* (pp. 97–124). New York, NY, US: Psychology Press.
Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335–384.
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.
Marsh, R. L., Meeks, J. T., Cook, G. I., Clark-Foos, A., Hicks, J. L., & Brewer, G. A. (2009). Retrieval constraints on the front end create differences in recollection on a subsequent test. *Journal of Memory and Language, 61*, 470–479.
McCabe, D. P., & Balota, D. A. (2007). Context effect on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 536–549.
McCabe, D. P., Geraci, L., Boman, J. K., Sensenig, A. E., & Rhodes, M. G. (2011). On the validity of remember–know judgments: Evidence from think aloud protocols. *Consciousness and Cognition, 20*, 1625–1633.
McCormack, P. D., & Swenson, A. L. (1972). Recognition memory for common and rare words. *Journal of Experimental Psychology, 95*, 72–77.
Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition, 21*, 1435–1455.
Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *The Quarterly Journal of Experimental Psychology, 62*, 1198–1215.
Reder, L. M., Nhouvanisvong, 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 the remember–know paradigm in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294–320.
Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*, 907–922.
Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43*, 67–88.
Rugg, M. D., Cox, C. J. C., Doyle, M. C., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia, 33*, 471–484.
Schmid, J., Herholz, S. C., Brandt, M., & Buchner, A. (2010). Recall-to-reject: The effect of category cues on false recognition. *Memory, 18*, 863–882.
Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*, 523–542.
Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia, 45*, 2163.
Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia, 47*, 1765–1779.
Steyvers, M., & Malmberg, K. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 760–766.
Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217.
Stretch, V., & Wixted, J. T. (1998). On the differences between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396.
Strong, E. K. Jr., (1912). The effect of length of series upon recognition memory. *Psychological Review, 19*, 447–462.
Tousignant, C., & Bodner, G. E. (2012). Test context affects recollection and familiarity ratings: Implications for measuring recognition experiences. *Consciousness and Cognition, 21*, 994–1000.
Underwood, B. J., & Freund, J. S. (1970). Word frequency and short-term recognition memory. *The American Journal of Psychology, 8*, 343–351.
Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20*, 400–405.
Williams, H. L., Conway, M. A., & Moulin, C. J. (2013). Remembering and knowing: Using another's subjective report to make inferences about memory strength and subjective experience. *Consciousness and Cognition, 22*, 572–588.
Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 681–690.
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.