Summer 8-15-2015

In Search of a Lost Effect: Generality of Discrepancy Effects in Memory Paradigms

Ji hae Lee
Washington University in St. Louis

Recommended Citation

Lee, Ji hae, "In Search of a Lost Effect: Generality of Discrepancy Effects in Memory Paradigms" (2015). *Arts & Sciences Electronic Theses and Dissertations*. 661.
https://openscholarship.wustl.edu/art_sci_etds/661

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
Mark McDaniel, Chair
David Balota
Janet Duchek
Henry Roediger, III
James Wertsch

In Search of a Lost Effect: Generality of Discrepancy Effects in Memory Paradigms

by

Ji hae Lee

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2015

Saint Louis, Missouri

Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgments

First, I would like to thank everyone on my committee, Dave, Jan, Jim, Mark, and Roddy, for their guidance and patience with this project. A special thanks goes to my advisor Mark McDaniel for his tremendous mentorship throughout this very challenging journey of graduate school. He set an example of how to be a collegial academic, showed how to find value in every set of data, and taught how not to give up on writing projects.

I thank former and current members of the McDaniel lab, the best lab any graduate student could wish for. I truly cherish all the stimulating discussions and encouraging smiles from the lab. Kwanjeong Educational Foundation built the foundation for all of my educational opportunities and the Buckley-Shaklee family provided a loving home over the last decade in the US. Khuyen Nguyen and Shannon Buckley-Shaklee served as my helpful editors on previous versions of my dissertation. I would also like to thank all my friends and mentors who helped me through this journey and whose names are not listed here in the interest of saving trees. However, their names are engraved in my heart.

Finally, I would like to thank my loving family, especially my parents, for their endless trust and support throughout my entire life.

마지막으로 저의 사랑하는 가족이, 그중에서도 저의 부모님이, 저의 평생에 보내준 끝없는 믿음과 도움에 감사를 표합니다.

Ji hae Lee
이지혜

Washington University in St. Louis
August 2015

ABSTRACT OF THE DISSERTATION

In Search of a Lost Effect: Generality of Discrepancy Effects in Memory Paradigms

by

Ji hae Lee

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2015

Professor Mark McDaniel, Chair

The current project investigated the generality of discrepancy effect in retrospective memory (RM) reported by Whittlesea and colleagues (e.g., Whittlesea & Williams, 2001a) and the generality of discrepancy effect in prospective memory (PM) reported by McDaniel and colleagues (e.g., McDaniel, Guynn, Einstein, & Breneiser, 2004). Experiments 1a and 2 tested the claim that discrepancy, elicited by mismatching the expected and the actual processing fluency, can give rise to familiarity under an RM context and increase familiarity judgments of discrepant items, independent of previous encounters with those items. Experiment 1b tested the claim that, within a PM context, such discrepancy can signal that discrepant items are significant and this significance can initiate the search for the source of the significance, thereby enhancing PM performance for discrepant PM cues. The current project attempted to elicit discrepancy by implementing a processing fluency paradigm with masked priming and a modified perceptual mask for Experiments 1a and 1b or high and low frequency words for Experiment 2. The discrepancy was manipulated by mismatching/matching the processing fluency of some items to the processing fluency of other items (e.g., fluent items embedded within disfluent items = discrepant items). In Experiment 1a, hit rates were higher for more fluently processed items (i.e., items with no perceptual mask) than less fluently processed items (items with a difficult perceptual mask), independent of discrepancy. In Experiment 2, hit rates were higher for low frequency words than high frequency words, independent of discrepancy. Furthermore, both in Experiments 1a and 2, false alarm rates did not differ as a function of discrepancy, fluency, or word frequency. In Experiment 1b, PM performance did not differ between discrepant and nondiscrepant PM cues. These results suggest that the discrepancy effects in RM and PM might not be as general as previously claimed.

Chapter 1: Introduction and Literature Review

How do we recognize faces of people we have previously encountered? How do we recognize material covered in the lecture while taking a multiple-choice exam? For decades, psychologists have investigated the basis of recognition judgment, the kind of information guiding the judgment of whether or not a particular item was previously encountered. One prevalent line of recognition research proposes that two distinctive processes influence recognition memory judgment, recollection and familiarity (e.g., Gardiner, 1988; Jacoby, 1983, 1991; Jacoby & Dallas, 1981; Mandler, 1980). Recollection refers to remembering something with detailed information about its previous encounter. Familiarity, on the other hand, refers to judging something as previously encountered based on the familiar feeling without the definite details of its previous encounter (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989; Whittlesea, 1993; for a review, see Yonelinas, 2002).

How do people remember to deliver a message to a colleague upon meeting him/her? How do people remember to turn off an oven while cooking? Over the years, researchers have investigated the basis of successful prospective memory (PM), the kind of information guiding one to remember to perform an intended action in the future. While a prompt to enter into a retrieval mode is provided for people when performing retrospective memory (RM) tasks (e.g., "Did you study this face?"), no such prompt is provided for PM tasks. For example, to perform a PM task of delivering a message to a colleague, one has to recognize the particular colleague as the PM cue while being busily engaged in a conversation with a group of colleagues and retrieve the PM intention associated with that cue without being prompted at the appropriate moment. One prevalent line of PM research proposes that multiple mechanisms underlie the recognition of the PM cue and the retrieval of the PM intention (Multiprocess framework, McDaniel & Einstein, 2000, 2007). According to the multiprocess framework, people may engage in strategic monitoring processes in some circumstances (e.g., when multiple PM cues are used, Cohen, Jaudas, & Gollwitzer, 2008) and rely on relatively spontaneous retrieval processes in other circumstances (e.g., single PM cue with a long delay between the encoding of PM intention and encountering of first PM cue, Scullin, McDaniel, Shelton, & Lee, 2010).

Although the explicit prompt for retrieval is present only in RM tasks, both RM and PM tasks might be supported by similar underlying processes. Some researchers have suggested that fluency (e.g., Jacoby & Dallas, 1981) or discrepancy (e.g., Whittlesea & Williams, 2001a) might give rise to familiarity in RM tasks and consequently influence recognition judgments. Other researchers have suggested that fluency (McDaniel, 1995) or discrepancy (e.g., McDaniel, Guynn, Einstein, & Breneiser, 2004) might give rise to significance in PM tasks and consequently influence the recognition of the PM cue and the retrieval of the PM intention. The current project investigated the role of discrepancy, as well as the role of fluency, in RM and PM tasks. Below, the theoretical roles of fluency and discrepancy for supporting familiarity processes in recognition memory tasks will be discussed first and then followed by their possible roles in PM tasks.

1.1 Familiarity Driven by Fluency

Researchers differ on what they think gives rise to familiarity (e.g., Jacoby & Dallas, 1981; Whittlesea & Williams, 1998). Jacoby and colleagues (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989) suggested that perceptual processing fluency of an item can give rise to familiarity and increase the likelihood of something to be judged as previously encountered, independent of its previous encounter (see Experiment 3 in Whittlesea, 1993, for evidence supporting that conceptual processing fluency of an item also can give rise to familiarity). They proposed that items with higher fluency in a recognition test are more likely to be judged as familiar than items with lower fluency (e.g., Jacoby, 1983; Jacoby, Kelley, & Dywan, 1989; Jacoby & Whitehouse, 1989; Johnston, Dark, & Jacoby, 1985; Kelley & Jacoby, 1998; Whittlesea, Jacoby, & Girard, 1990). In support of this proposal, Jacoby and colleagues have shown that perceptually more fluently processed items in a recognition test were more likely judged as having been previously encountered than perceptually less fluently processed items, whether those items were previously encountered (correct "old" response to a recognition task termed as *hit*) or not encountered (incorrect "old" response termed as *false alarm*).

For example, Whittlesea, Jacoby, and Girard (1990) manipulated the perceptual fluency of words during a recognition memory test by altering the density of dots that composed the perceptual mask covering both studied and nonstudied words. They observed that previously studied words were more likely judged to have been studied than those that were not previously studied. Moreover, they found that recognition probes with the lower density perceptual mask were identified and processed more fluently, based on the faster reaction times (RTs) of pronunciation of the probes. More importantly, these probes were more likely judged as having been previously studied than probes with the higher density perceptual mask that were processed less fluently, whether the recognition probes were previously studied or not. Another important aspect of this effect is that only when participants were unaware of the source of varying fluency (i.e., density of perceptual mask), more fluently processed items received higher rates of hit and false alarm (in Experiments 1 and 2). On the other hand, when participants were told of the source of varying fluency, higher fluency did not increase the familiarity judgments for recognition probes, whether they were studied or not (in Experiment 3). Whittlesea et al. interpreted their findings to indicate that the processing fluency of an item during the recognition test can be interpreted and attributed as familiarity, and this incorrect attribution is possible when participants do not take the correct source of fluency into consideration. Using processing fluency as a basis for familiarity is an efficient heuristic because previously encountered, familiar items are processed more fluently than previously not encountered, unfamiliar items. When participants can identify the correct source of increased fluency, they attribute the fluency to its correct source and consequently do not misattribute the fluency to familiarity.

Another example showing that processing fluency influences familiarity judgments is with the use of a "context" word prior to the presentation of recognition probes. During a recognition memory test of words, Jacoby and Whitehouse (1989) presented a word that was either identical or unrelated to recognition probes for a brief 50 or long 200 msec in Experiment 1 (or 16 versus 600 msec in Experiment 2). Their reasoning was that the presentation of an identical "context" word can facilitate the processing of a subsequent recognition probe compared to the presentation of a no "context" word. Presentation of an unrelated "context" word, however, can disrupt the processing of a subsequent recognition probe compared to the presentation of a no "context" word. Furthermore, with the short presentation duration (e.g., 50 or 16 msec) of "context" words, participants would be unable to attribute the relatively higher fluency to its correct source, thereby falsely attributing the heightened fluency to familiarity. In support of this reasoning, with the 50 msec of a "context" word presentation, both studied and nonstudied recognition probes preceded by identical “context” words were more likely judged as previously studied compared to the baseline items with no “context” words. Recognition probes preceded by unrelated “context” words were less likely judged as previously studied compared to the baseline items. With the long presentation duration (e.g., 200 or 600 msec), participants could correctly identify the source of the increased fluency of recognition probes. This led participants to discount the fluency of recognition probes preceded by the identical “context” words, lowering the likelihood of judging those probes as previously studied, either the probes were studied or nonstudied, compared to the probes preceded by unrelated or no “context” words.

1.2 Familiarity Driven by Discrepancy

Despite the fact that the fluency-driven familiarity view (*fluency attribution account*) has been able to explain findings in the recognition memory literature (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989; Whittlesea et al., 1990; Whittlesea, 1993), Whittlesea and colleagues (Whittlesea & Williams, 1998, 2000, 2001a, 2001b) have proposed that it is not fluency per se that gives rise to familiarity. Instead, according to Whittlesea and colleagues, discrepancy that is induced by the violation of expected and experienced processing fluency gives rise to familiarity. In support of this proposal, Whittlesea and Williams (1998, 2000, 2001a, 2001b) have shown in various experiments that less fluently processed items can be more likely judged as previously studied than more fluently processed items. Based on these findings, Whittlesea and Williams (e.g., 2001a) developed the discrepancy attribution account. According to this account, people chronically evaluate the quality of their mental processing by comparing the expected processing quality and the actual processing quality. The expected processing quality is constructed from one’s knowledge or previous experience, whereas the actual processing quality is developed from the online experience. If the expected quality matches the experienced quality, no discrepancy is signaled. However, if they mismatch, discrepancy is signaled. Once discrepancy is signaled, the cognitive system attempts to resolve that discrepancy by attributing it to a plausible, but not necessarily correct, source. In the context of a recognition memory test, discrepancy of recognition probes can be attributed as familiarity of those probes and consequently increase the endorsement of “old” responses for them. Whittlesea and colleagues found substantial support for their framework in the form of higher false alarm and hit rates for the discrepant items compared to the nondiscrepant items (e.g., Whittlesea & Williams, 1998, 2000, 2001a, 2001b).

In Experiments 1 and 2 (1998), Whittlesea and Williams had their participants study a list of words, pronounceable nonwords (e.g., BELINT, LAFER), and pseudo-homophones that sounded the same as regular English words but had unusual spelling (e.g., PHRAWG, KANSER). Prior to making a recognition judgment for each test probe, participants were asked to pronounce the test probe (and additionally make a lexical decision on it in some, but not all, experiments), which emphasized the difference in processing fluency across words, pronounceable nonwords, and pseudo-homophones. If fluency were to guide familiarity judgments, false alarm rates should have been the highest for the words with the fastest pronunciation (and lexical decision) RTs, followed by the pseudo-homophones which were pronounced faster than pronounceable nonwords.

However, Whittlesea and Williams (1998) found higher false alarm rates, not for the words, but for the pseudo-homophones that had slower RTs than words. According to Whittlesea and Williams, participants might initially develop an expectation of less fluent processing of a pseudo-homophone because of its unusual spelling. However, when participants experience a fluent reminding of a real word that has the same sound as the pseudo-homophone, they might find that experience surprising and this surprise could elicit discrepancy. This discrepancy could increase the false alarm rate of pseudo-homophones relative to that of words or pronounceable nonwords.

Although the hit rate of pseudo-homophones was comparable to that of words, Whittlesea and Williams (1998) argued that the comparison of hit rates between different types of recognition probes in such an experiment could be misleading. Whittlesea and Williams suggested that memorability of previously studied probes (e.g., ease with elaborate encoding) could vary between words and nonwords and subsequently lead to different hit rates of words and nonwords even in the absence of discrepancy. On the other hand, memorability for nonstudied probes should not vary as much at the baseline level, allowing an easier observation of the discrepancy effect. They further argued that fluent processing of words and less fluent processing of pronounceable nonwords simply matched the expected processing fluency for each stimulus type, therefore did not elicit discrepancy, and consequently did not increase familiarity and false alarm rates for those stimulus types.

In another series of experiments, Whittlesea and colleague (Whittlesea & Williams, 1998, 2000, 2001a) had their participants study a list of words (e.g., DAISY, RAINBOW), orthographically regular nonwords (e.g., HENSION, PINGLE), and orthographically irregular nonwords (e.g., STOFWUS, LICTPUB). Again, if fluency was to guide familiarity judgments, (hit and) false alarm rates should have been the highest for the words with the fastest pronunciation and lexical decision RTs, followed by the orthographically regular nonwords which were processed faster than orthographically irregular nonwords. However, Whittlesea and Williams found higher false alarm rates for the orthographically regular nonwords compared to both words and orthographically irregular nonwords. Their explanation for this finding was that surprisingly more fluent processing of orthographically regular nonwords violated one’s expectation for processing of nonwords and subsequently elicited discrepancy. This discrepancy was incorrectly attributed as familiarity originating from a previous study episode because participants failed to take orthographical regularity into consideration when attributing fluency to its possible source.

In addition to reporting experiments that showed the discrepancy-enhanced familiarity judgments described above, Whittlesea and Williams reinterpreted the data that were previously explained by fluency as being explained by discrepancy. Whittlesea and Williams (Whittlesea & Williams, 2001b; see also, Whittlesea, 1993) used a set of words that served as the final word in various types of sentences. After studying a list of words, participants were asked to read a sentence without the final word and make a recognition judgment on the final word upon its subsequent presentation. Some of the sentences were written such that the final word (with several other words) can be semantically consistent with the rest of the sentence (e.g., “Broom” as the final word for “She couldn't find a place to put the ...”). Other sentences were written such that the final word was (somewhat but not definitely) predictive from the rest of the sentence (e.g., “Beach” for “They swam and played at the …”). According to the fluency attribution account, final words presented in a predictive sentence should be judged more familiar than those in a consistent sentence given that the former are read faster than the latter (e.g., Experiment 2 in Whittlesea, 1993). Consistent with this prediction, Whittlesea (Experiment 3, 1993) found higher hit and false alarm rates for final words presented in the predictive sentence compared to those in the consistent sentence.

However, in later work, Whittlesea and Williams (2001b) provided a revised interpretation of the higher hit and false alarm rates for final words presented in the predictive sentences reported in the previous work (Whittlesea, 1993). They argued that such data were driven by discrepancy, instead of fluency, that was induced by interaction of sentence type and the design feature of how the final words were presented. They argued that the 2-second presentation rate of a sentence stem prior to the presentation of a final word could have allowed participants to finish reading the sentence stem and occasionally experience a pause prior to the presentation of a final word. This occasional pause does not induce discrepancy for the final words in the consistent sentence stems because participants would not expect any particular final word to follow. However, this occasional pause in the predictive sentence stems could induce discrepancy because predictive-sentence stems lead participants to build general expectations for possible final words and to experience uncertainty with the pause. When the final word is presented after the pause, participants might find the final word to fit the predictive sentence surprisingly well and attribute that surprise, discrepancy, to familiarity.

To test this revised interpretation, Whittlesea and Williams (2001b) experimentally manipulated the presence of a pause between different types of sentence stems and final words. Indeed, Whittlesea and Williams found higher false alarm rates for final words in the predictive sentence stem compared to those in the consistent sentence stem when the pause (250 msec) preceded the final word. They also found higher hit rates for the discrepant final words (that were presented after a pause in the predictive sentence stem) than for the nondiscrepant final words. They concluded that when the presence and absence of discrepancy is manipulated within the same stimulus type (e.g., words serving as either discrepant or nondiscrepant items), one can observe the effect of discrepancy on hit rates as well as on false alarm rates. According to Whittlesea and Williams, this comparison (of hit rates) was not informative in experiments where the presence and absence of discrepancy were manipulated across different stimulus types (e.g., orthographically regular nonwords and words serving as discrepant and nondiscrepant items, respectively). Rates of false alarms and hits did not differ as a function of sentence stem type when no pause was inserted prior to the presentation of final words, corroborating the discrepancy-based interpretation over the fluency-based interpretation.

Whittlesea and colleagues have argued that discrepancy is a robust effect by showing its influence across a range of conditions including recognition memory (e.g., Whittlesea & Leboe, 2003; Whittlesea & Williams, 1998, 2000, 2001a, 2001b; but see Cleary et al., 2007), false memory (Whittlesea, 2002; Whittlesea, Masson, & Hughes, 2005; but see Karpicke, McCabe, & Roediger, 2008), the revelation effect (e.g., Bernstein, Whittlesea & Loftus, 2002) and various types of materials (e.g., words and nonwords, Whittlesea & Williams, 1998; words in different sentences, Whittlesea & Williams, 2001b; and musical tones, Whittlesea & Williams, 2001a). Furthermore, although discrepancy is attributed as familiarity of items in the context of a recognition memory test, they suggested that it can also be attributed as something else in different contexts (e.g., Whittlesea & Williams, 2001a). In support of this argument, studies have found that discrepancy can influence various judgments, such as preference (Willems, Van der Linden, & Bastin, 2007) or subjective truth (Hansen, Dechêne, & Wänke, 2008).

For example, Willems et al. (2007) had their participants study a mix of clear and blurred pictures of faces and make either recognition judgments with remember/know responses or preference judgments on those faces. They argued that participants would build an expectation for lower processing fluency for blurred pictures compared to clear pictures.

However, when both types of pictures were previously studied and the study episode increased fluency for those pictures (compared to those that were not previously studied), only the enhanced fluency of the blurred pictures, not that of clear pictures, would violate that expectation and induce discrepancy. In line with this idea, Willems et al. found that participants made more “know” judgments, which are driven by familiarity, for previously studied blurred pictures than previously studied clear pictures. Furthermore, “remember” judgments, which are driven by recollection, were the same between the two types of pictures, providing additional support for the discrepancy attribution account. Moreover, when the participants were asked to make preference judgments on the half of studied pictures that were not used for the recognition memory task, they judged the blurred pictures more preferable than the clear pictures. Based on these findings, Willems et al. argued that the same discrepancy that was attributed as familiarity for the blurred pictures in the context of recognition memory task could be attributed as preference in the context of a preference task.

Hansen et al. (2008) also provided additional evidence supporting the claim that discrepancy can be attributed as something other than familiarity. Their participants rated the subjective truth of a list of sentences that were written either with a greater color contrast between font and background, thus, perceptually fluent, or with a lower color contrast, thus, disfluent. Fluent sentences were rated truer when they were preceded by disfluent sentences, thus, discrepant, than when they were preceded by fluent sentences, thus, nondiscrepant. Findings from Hansen et al. as well as Willems et al. (2007) support the claim that discrepancy can be attributed to a wide range of plausible sources.

1.3 Discrepancy Effects on Prospective Memory

The contextual sensitivity of discrepancy attribution reviewed above encouraged

McDaniel and his colleagues to adopt discrepancy as a possible mechanism underlying PM (the discrepancy-plus-search account, McDaniel et al., 2004). The typical laboratory event-based PM task requires participants to engage in an ongoing activity, such as a lexical decision task (LDT) where a string of letters is judged either as a word or nonword. In addition to the ongoing task demand, participants are instructed to make a PM response (e.g., pressing the “q” key) whenever they see a particular PM cue appear (e.g., any word starting with the letter “o”) during the ongoing task. To make the correct PM response, participants have to recognize a stimulus as the PM cue that is associated with the PM intention, while engaged in the ongoing task. Once a stimulus is recognized as the PM cue, participants have to retrieve the PM intention associated with the PM cue. Given that the recognition and retrieval in a PM task have to be initiated without being explicitly prompted, McDaniel et al. suggested that, within the PM context, discrepancy from a stimulus, elicited by a mismatch between the expected processing quality and the actual processing quality, might be interpreted as indicating significance (rather than familiarity) of that stimulus. The significance of the stimulus then is assumed to serve as an exogenous cue, promoting the search for the source of that significance. This search likely leads to the recognition of the item as a PM cue and the retrieval of the PM intention.

Studies have found support for the discrepancy-plus-search account (Breneiser & McDaniel, 2006; Guynn & McDaniel, 2007; Lee & McDaniel, 2013; McDaniel et al., 2004; Thomas & McBride, 2015). For instance, in a paradigm where the ongoing task was to solve anagrams and the PM task was to press a specific key for anagrams of particular words (e.g., anagrams for “lawyer” or “orange”), Lee and McDaniel manipulated discrepancy by mismatching the expected and the actual difficulty level of the anagram solution for PM cues.

In this experiment, the use of subliminal priming of the anagram solution and the varying degrees of letter dislocation in anagrams allowed the experimenters to vary the difficulty of anagrams, hence, processing fluency of them, without participants being able to specify the source of fluency. The expectation (of subsequent anagrams' difficulty) was built by having participants solve a list of anagrams with a particular solution difficulty. For example, those who solved a list of anagrams with easy solutions presumably developed the expectation that the subsequent anagrams would be easy to solve. This expectation was met by PM cue anagrams with easy solutions in the nondiscrepant condition and was violated by PM cue anagrams with difficult solutions in the discrepant condition. In the other conditions, participants developed the expectation for anagrams with difficult solutions in both the nondiscrepant and the discrepant conditions, respectively.

Lee and McDaniel (2013) found their participants were more likely to make the PM response of pressing the "q" key while solving anagrams for the discrepant PM cue anagrams than for the nondiscrepant PM cue anagrams. More specifically, both easy PM cues embedded in the difficult list and difficult PM cues embedded in the easy list showed the PM improvement. Furthermore, lack of relative slowing of RTs for nontarget anagrams in the PM block compared to that in the control block corroborated that the higher PM performance for the discrepant PM cues were more likely to be driven by discrepancy processes. If monitoring processes enhanced PM performance for the discrepant PM cues, the RTs of nontarget trials in the PM block would have been slower compared to the RTs of nontarget trials in the control block (McDaniel et al., 2004; Smith, 2003). Lee and McDaniel's findings suggest that discrepancy can indeed facilitate PM performance. If fluency of PM cue, not discrepancy, enhanced PM performance according to the familiarity view proposed by McDaniel (1995), a main effect of PM cue difficulty would have been observed. However, it is noteworthy that the highest PM performance in their experiment was found not with the easy PM cues but with the difficult PM cues embedded in the list of easy anagrams. Corroborating the findings of Lee and McDaniel, Thomas and McBride (2015) also found higher PM performance for discrepant PM cues that were processed less fluently due to semantic incongruence with the rest of ongoing task stimuli (i.e., PM cues being exemplars of less dominant category during a category decision task) compared to nondiscrepant PM cues that were fluently processed (PM cues being exemplars of more dominant category).

1.4 Counter-arguments for Discrepancy Effects in Retrospective Memory

Although studies have provided support for the claims stating that discrepancy can give rise to familiarity (e.g., Goldinger & Hansen, 2005; Whittlesea & Williams, 1998, 2000, 2001a, 2001b) and can be attributed to something other than familiarity across different contexts (Lee & McDaniel, 2013; Hansen et al., 2008; Thomas & McBride, 2015; Willems et al., 2007), discrepancy might not be as potent a phenomenon as Whittlesea and colleagues have claimed for a number of reasons. First, most of the items with which Whittlesea claimed to have found the discrepancy effect could be considered as relatively more fluent items, instead of relatively disfluent items. For example, orthographically regular nonwords are pronounced less fluently than words, but are pronounced more fluently than orthographically irregular nonwords (e.g., Whittlesea & Williams, 1998). If so, the fluency attribution account can provide an explanation for increased familiarity judgments for those items. Also, several studies have provided alternative explanations for the discrepancy effect on false memory (Karpicke et al., 2008) and recognition memory (Clearly et al., 2007).

Consider, for example, the Deese–Roediger–McDermott (DRM) effect. Whittlesea and colleagues (Whittlesea, 2002; Whittlesea et al., 2005) argued that discrepancy might underlie *false memory*. False memory refers to increased false recognition or recall of critical items (that are not presented) that are strongly associated with a list of items that was studied compared to critical items of a nonstudied list (e.g., Roediger & McDermott, 1995). More specifically, Whittlesea (2002) claimed that although the studying of associates of critical items enhances semantic processing of critical items, the lack of study of critical items would lead to perceptual processing of critical items being surprisingly not fluent, hence, inducing discrepancy. Subsequently, this surprise-induced discrepancy can be attributed as familiarity of critical items and increase false memory for them. However, Karpicke et al. (2008) tested and disputed this claim by directly asking their participants if they experienced surprise with any recognition probes during a memory test. They found typical false memory for critical items of the studied list but did not find those items to be judged more surprising (in Experiment 1) or less readable (Experiment 4) compared to critical items of the nonstudied list. Based on these findings, Karpicke et al. ruled out Whittlesea's claim that surprise was experienced for the critical items as well as the claim that the surprise led to discrepancy for those items and increased false memory for them.

Cleary et al. (2007) also discounted the discrepancy attribution account using the structural regularity hypothesis. According to their structural regularity hypothesis, knowledge of structural regularity and reliance on such knowledge during learning might enhance learning of new information. Based on this hypothesis, orthographically regular nonwords should be better remembered than orthographically irregular nonwords. Thus, the higher false alarm rate of the orthographically regular nonwords compared to that of the orthographically irregular nonwords (Whittlesea & Williams, 1998, 2000, 2001a) is inconsistent with the structural regularity hypothesis.

Cleary et al. (2007) conducted a series of experiments with a set of stimuli they constructed using strings of words (Experiment 1a) and line drawings (Experiment 1b) and stimuli from Whittlesea and Williams (1998; i.e., words, and orthographically regular versus irregular nonwords for Experiment 1c) as recognition probes. Cleary et al. were able to replicate findings from Whittlesea and Williams (1998, 2000, 2001a), showing the higher false alarm rate (as well as higher hit rate) of orthographically regular nonwords compared to that of orthographically irregular nonwords or that of words only in selective experiments. These experiments used the same materials and procedure from Whittlesea and Williams (Experiment 2) or the modified procedures, without either the pronunciation task or the LDT on the recognition probe prior to the recognition judgment (Experiments 3a and 3b, respectively). These replications corroborate the discrepancy attribution account.

However, in other experiments, Cleary et al. (2007) failed to replicate the higher false alarm rate of orthographically regular nonwords compared to that of irregular nonwords, although they were able to replicate the higher false alarm rate of orthographically regular nonwords to that of words. Cleary et al. emphasized the importance of the comparison of orthographically regular nonwords and irregular nonwords for a number of reasons. They argued that, although the comparison of orthographically regular nonwords and words has been reported repeatedly (e.g., Greene, 2004; Whittlesea & Williams, 2000, 2001a, 2001b), the comparison of orthographically regular and irregular nonwords has not. Furthermore, they suggested that the higher false alarm rate of orthographically regular nonwords than that of words is consistent with the pseudoword effect (Greene, 2004) and does not argue against the structural regularity hypothesis.

Cleary et al. (2007) found comparable rates of false alarms for the items that functionally correspond to the categories of “orthographically regular nonwords” and “orthographically irregular nonwords” in a number of experiments. Those experiments had recognition probes that were strings of words (Experiment 1a) or line drawings (Experiment 1b) or stimuli from Whittlesea and Williams (1998) with procedural modifications (Experiment 1c). They also failed to find an effect of discrepancy when participants were not asked to perform either the pronunciation task or the LDT prior to the recognition judgment (Experiment 3c) or were asked to engage in a secondary articulatory suppression task during the recognition test (Experiment 3d). Instead, in those experiments, they found no difference on false alarm rates between regular nonwords (by their definition, meaningless items with structural regularity) and irregular nonwords (meaningless items without structural regularity). They also found higher hit rates for regular nonwords compared to that of irregular nonwords. They interpreted these findings as evidence supporting the structural regularity hypothesis.

Based on their findings, Cleary et al. (2007) argued that the discrepancy effect might be driven by phonological factors and confounded with higher inter-stimulus similarity of nonstudied orthographically regular nonwords to the rest of studied stimuli compared to nonstudied orthographically irregular nonwords or words. Higher inter-stimulus similarity of nonstudied orthographically regular nonwords to the rest of studied stimuli has been found to increase false memory (Westbury, Buchanan, & Brown, 2002). Indeed, when inter-stimulus similarity was controlled, they found lower false alarm rates for nonwords with a higher number of orthographic neighbors compared to nonwords with a lower number of orthographic neighbors and comparable hit rates between the two types of nonwords (Experiment 5). Again, these patterns of results suggest better old-new discrimination for regular nonwords than irregular nonwords, which supports the structural regularity hypothesis and discounts the discrepancy attribution account.

Given the opposing patterns of data and their interpretations, the generality of discrepancy in RM seems undetermined. For example, even though Cleary et al. (2007) discounted the generality of the discrepancy effect with the orthographically regular nonwords, they admitted that their findings and interpretations do not account for the mechanism underlying the discrepancy effect in the sentence-with-a-pause paradigm (Whittlesea & Williams, 2001b). Thus, one aim of the current project is to test the generality of discrepancy in RM with materials and paradigms not used previously.

1.5 Counter-arguments for Discrepancy Effects in Prospective Memory

The role of discrepancy in PM can also be questioned. Though suggestive, the small number of existing studies on discrepancy and PM are subject to criticisms that discrepancy is not the only explanation for the results (because discrepancy was induced by high familiarity of PM cues relative to nontargets, Guynn & McDaniel, 2007) or that the attempts to manipulate discrepancy in the PM task are not comparable to that discussed by Whittlesea and colleagues (because participants could have been aware of the source of discrepancy, Breneiser & McDaniel, 2006). Currently, Lee and McDaniel's (2013) finding is the most convincing evidence suggesting that, within a PM context, discrepancy might be attributed as significance, thereby leading to the search for the source of that significance and enhancing PM performance. However, that experiment examined the effect of discrepancy only on PM performance. The discrepancy attribution account claims that the same discrepancy can lead to different attributions in different contexts (Whittlesea & Williams, 2001a), yet no project has investigated if the identical or even comparable manipulations of discrepancy can influence both RM and PM. Thus, a second objective of the current project is to test if a similar manipulation of discrepancy can lead to different attributions (e.g., familiarity in an RM task context and significance in a PM task context), thereby influencing both RM and PM performance.

Chapter 2: Overview of Experiments of the Current Project

A number of factors pose challenges to designing a convincing experimental paradigm that can address the objectives stated above. The challenges arise because paradigms that tested discrepancy in RM often induced discrepancy by manipulating (1) stimulus characteristics (2) in very specific settings, both of which are difficult to incorporate into a PM task. Whittlesea and Williams (1998, 2001a) argued that they elicited discrepancy for orthographically regular nonwords and pseudo-homophones and that discrepancy increased false alarms for orthographically regular nonwords and pseudo-homophones compared to words. The former types of items, along with words, could be used as PM cues. However, even if the PM cues of the former types exhibit higher PM performance than the PM cues of words, an interpretational problem would arise. That higher PM performance could be caused by discrepancy, which presumably increased false alarm rates in RM, or by a factor other than discrepancy. For instance, different characteristics of orthographically regular nonwords compared to that of words could lead to more elaborate encoding of orthographically regular nonwords. Participants might attempt a more elaborate encoding of orthographically regular nonwords because the relatively less-fluent processing of those items, compared to that of words, could lead participants to perceive that searching for PM cues of orthographically regular nonwords is more challenging. This interpretational problem limits the incorporation of Whittlesea's paradigms that elicited discrepancy for items with specific characteristics into a PM task.

Another way to induce discrepancy is via procedural techniques. For example, Whittlesea and Williams (2001b) induced discrepancy by manipulating the presence of a pause inserted between words within a sentence with varying levels of semantic predictability. Their manipulation led to results consistent with the discrepancy attribution account: words with discrepancy elicited by procedural techniques showed higher hit and false alarm rates than words with no discrepancy. Such a manipulation of discrepancy could eliminate the interpretational problem discussed above. However, when researchers attempted to utilize this paradigm with a PM task (Lee & McDaniel, unpublished data), they were faced with another interpretational problem. The presence of a pause, that Whittlesea and Williams used to manipulate discrepancy, created a confound. Specifically, the pause could have allowed a longer time for participants to process PM cues in the discrepant condition than in the nondiscrepant condition (that had no pause). If so, the pause could have subsequently enhanced PM performance in the discrepancy condition by allowing more processing time (or monitoring) of PM cues rather than by inducing discrepancy per se.

2.1 Overview of Experiments 1a and 1b

However, other procedural techniques that manipulate processing fluency could be used to elicit discrepancy without the interpretational problems discussed above. Such techniques have been used to induce discrepancy in a PM paradigm (e.g., masked priming, Lee & McDaniel, 2013). Based on these findings, I used a paradigm that manipulated processing fluency in order to create discrepancy. More specifically, I manipulated processing fluency of a particular set of test probes, either easy or difficult, as well as processing fluency of other probes within which the particular set was embedded, either easy or difficult. Doing so led to some conditions having fluency consistent across test probes (all easy or all difficult), thus, having no discrepancy, whereas other conditions having fluency inconsistent across test probes (some easy and some difficult), thus, having discrepancy.

To test the generality of the discrepancy effect in RM (Experiment 1a) and the generality of the discrepancy effect in PM (Experiment 1b), I adapted a processing fluency paradigm with a perceptual mask from Whittlesea et al. (1990). Whittlesea et al. used a perceptual mask (i.e., layer of dots with a varying density) to manipulate processing fluency of test probes and showed that more fluently processed test probes were judged with a higher familiarity than less fluently processed probes. Therefore, I also used a perceptual mask (i.e., a string of symbols) to manipulate processing fluency in the current project. Some test probes were covered with a perceptual mask that made the identification of the probes relatively difficult whereas other test probes were not covered with a perceptual mask which made the identification relatively easy. For example, to make the identification difficult, for a probe presented in black font on a white background, a perceptual mask composed of a string of white symbols (@#$%&*?8) was laid over the probe. Doing so led the difficult probes to look as if some parts of the probe in black were erased (because of the coverage by white dots). For the easy probes, this layer of white symbols was absent. To make the presence of perceptual mask less obvious, both the easy and difficult probes were covered by a layer of colored (e.g., red) symbols (@#$%&*?8).

In addition to the modified perceptual mask, the current project implemented the masked priming to manipulate processing fluency. Extensive research exists showing the effect of masked priming on processing fluency (e.g., Forster & Davis, 1984; Rajaram & Neely, 1992; Weldon, 1991). The advantage with the use of masked priming is that it has been found to enhance processing fluency without participants being aware of the source of that fluency, which is a critical component for both the fluency attribution account and the discrepancy attribution account. As described previously, both views argue that only the enhanced fluency (by the fluency attribution account) or discrepant fluency (by the discrepancy attribution account) of its source being unidentified can be falsely attributed as familiarity. In the current project, test probes paired with no perceptual mask were primed with identity primes to further facilitate the processing of the probes whereas test probes paired with the perceptual mask (of white symbols) were primed with disrupting primes to further hinder the processing of the probes.

Combining those two techniques allowed Experiment 1a to investigate the role of fluency by comparing RM performance on the easy probes (with no perceptual mask and the identity prime) to the difficult probes (with the difficult perceptual mask and the disrupting prime). Also, such techniques allowed for the investigation of the role of discrepancy in PM. The easy (or difficult) probes embedded among the difficult (or easy) probes were considered as the discrepant probes whereas the easy (or difficult) probes embedded among the easy (or difficult) probes were considered as the nondiscrepant probes.

According to the discrepancy attribution account (Whittlesea & Williams, 1998, 2001a), discrepancy can give rise to familiarity in a recognition task context and that familiarity can increase the likelihood of responding “old” on a recognition judgment of discrepant items. Whittlesea and colleagues mostly focused on the analysis of false alarm rates when observing the discrepancy effect in paradigms that manipulated discrepancy by stimulus characteristics because, in such paradigms, baseline hit rates differed between items that were discrepant and nondiscrepant, making the analysis of hit rates less informative (e.g., Whittlesea & Williams, 1998). Indeed, when discrepancy was manipulated by factors other than characteristics of recognition probes, such as predictability of a sentence stem and the presence of pause (Whittlesea & Williams, 2001b), the discrepancy increased the rates of both hits and false alarms. Given that Experiment 1a manipulated discrepancy as a function of perceptual mask and masked priming (not by stimulus characteristics), the discrepancy effect is expected to be observed on both hit and false alarm rates in the current project. More specifically, higher false alarm and hit rates should be observed on the discrepant probes compared to the nondiscrepant probes for experiments that test RM (Experiment 1a). By contrast, according to the fluency attribution account, higher false alarm and hit rates should be observed on the easy probes compared to the difficult probes.

According to the discrepancy-plus-search account (McDaniel et al., 2004; Lee & McDaniel., 2013), discrepancy can give rise to significance, instead of familiarity, for discrepant items in a PM task context. This significance then serves as an exogenous cue, initiating the search for the source of significance. The search consequently increases likelihood of the recognition of PM cue and the retrieval of PM intention. Thus, in an experiment that tests PM performance (Experiment 1b) higher PM performance is expected for the discrepant PM cues compared to the nondiscrepant PM cues.

2.2 Overview of Experiment 2

Experiment 2 also tested the generality of the discrepancy effect in RM. Instead of using the modified perceptual mask used for Experiments 1a and 1b, Experiment 2 used high and low frequency words, in addition to the masked priming, to manipulate processing fluency. The reasoning was that participants would have different expectations for how fluently high and low frequency words should be processed, based on their pre-experimental experience with those words. These expectations can be met or violated by the implementation of facilitative or disruptive primes. For example, participants might find the increase in processing fluency by an identity prime more discrepant than the decrease in processing fluency by a disrupting prime for low frequency words because they expect those words to be processed less fluently. For high frequency words, on the other hand, participants might find the decrease in processing fluency by a disrupting prime more discrepant than the increase in processing fluency by an identity prime because they expect those words to be processed more fluently.

Thus, in Experiment 2, low frequency words with the identity prime and high frequency words with the disrupting prime should show higher false alarm rates than low frequency words with the disrupting prime and high frequency words with the identity prime, respectively. Furthermore, considering that combining high and low frequency words with the masked priming has an advantage of allowing discrepancy to be present or absent within the same stimulus type (e.g., low frequency words being either discrepant or nondiscrepant depending on the type of primes), the discrepancy effect is expected to be observed also on hit rates. More specifically, higher hit rates should be observed on low frequency words with the identity prime and high frequency words with the disrupting prime than on low frequency words with the disrupting prime and high frequency words with the identity prime, respectively. By contrast, according to the fluency attribution account, higher false alarm and hit rates should be observed on the probes with the identity prime compared to the probes with the disrupting prime, independent of word frequency (see, however, Kinoshita, 1995, for evidence of greater repetition priming for low frequency words relative to high frequency words).

Chapter 3: Experiment 1a

There are more experiments examining the effect of discrepancy on RM (e.g., Cleary et al., 2007; Whittlesea, 1993; Whittlesea & Williams, 1998, 2000, 2001a, 2001b) than on PM (Breneiser & McDaniel, 2006; Guynn & McDaniel, 2007; Lee & McDaniel, 2013; McDaniel et al., 2004; Thomas & McBride, 2015). Experiments 1a and 1b attempted to demonstrate that the proposed paradigm can induce discrepancy by mismatching the expected processing fluency to the experienced processing fluency. Experiment 1a tested if such induction of discrepancy can give rise to familiarity and thus influence RM performance of discrepant items, whereas Experiment 1b tested if such discrepancy can give rise to significance and thus influence PM performance of discrepant items.

In Experiment 1a, participants studied a list of words and nonwords. During the test phase, participants were asked to make a recognition judgment on a test probe. The perceptual processing fluency of studied and nonstudied test probes was manipulated by utilizing a perceptual mask and masked prime. There were two levels of processing fluency of test probes. Easy probes were paired with no perceptual mask and the identity prime whereas difficult probes were paired with the difficult perceptual mask and the disrupting prime. For some groups of participants, the processing fluency of a particular set of test probes (hereafter referred to as *critical item* difficulty; easy or difficult) differed from the processing fluency of the rest of the test probes (hereafter referred to as *noncritical item* difficulty; difficult or easy, respectively). These groups were considered to be in discrepant conditions because the mismatching processing fluency of critical items, compared to that of the noncritical items, is supposed to elicit discrepancy for the critical items. In other groups, all test items, both the critical items and the noncritical items, were of the same processing fluency, either easy or difficult. These groups were considered to be in nondiscrepant conditions because the matching processing fluency of critical items to that of noncritical items is supposed to elicit no discrepancy for the critical items.

Critical items were every 17th and 18th of test probes preceded by 16 noncritical items. Such a design feature was implemented to encourage participants to build an expectation of particular processing fluency for critical items after performing a recognition task for a list of noncritical items with a particular processing fluency. For example, those who made recognition judgments on a list of test probes (16 noncritical items) that were perceptually fluent presumably would develop the expectation that the subsequent test probes would be easy to process. This expectation would be met by easy critical items, eliciting no discrepancy, or would be violated by difficult critical items, eliciting discrepancy. When the expected processing fluency is difficult, difficult critical items would elicit no discrepancy and easy critical items would elicit discrepancy.

Whittlesea and colleagues (e.g., Whittlesea & William, 1998) implemented a LDT or a pronunciation task prior to a recognition judgment of a test probe to emphasize the processing fluency of test probes in paradigms that tested effects of discrepancy on familiarity judgments. Thus, Experiment 1 also implemented a LDT prior to making a recognition judgment of each test probe. Implementation of a LDT in Experiment 1a would emphasize the processing fluency manipulation of test probes. Furthermore, the implementation of a LDT would provide measures to test the efficacy of the processing fluency manipulation, independent of whether or not discrepancy can affect recognition performance of test probes. Successful manipulation of processing fluency with the perceptual mask and masked priming would lead to faster RTs and higher accuracy for the easy probes compared to the difficult probes in the LDT performed immediately prior to making a recognition judgment.

With the processing fluency manipulation, discrepancy should occur for the critical items with mismatching processing fluency (i.e., easy probes embedded within a list of difficult probes and difficult probes embedded within a list of easy probes) compared to that of the noncritical items, instead of for the critical items with matching processing fluency (i.e., easy probes embedded within a list of easy probes and difficult probes embedded within a list of difficult probes). If discrepancy gives rise to familiarity in the recognition context, then critical items with discrepancy should be experienced as more familiar than critical items with (relatively) no discrepancy. Such enhancement in familiarity of critical items with discrepancy should increase the endorsement of them as a previously presented item, thereby increasing hits for studied critical items and false alarms for nonstudied critical items that were in the discrepant conditions compared to those in the nondiscrepant conditions. In other words, if the effect of discrepancy was to be observed, I would find a crossover interaction of critical item difficulty and list difficulty (which refers to the difficulty of noncritical items) on hit and false alarm rates of critical items (see Figure 1 for the predicted pattern of results).

By contrast, if fluency were the underlying mechanism that gives rise to familiarity, a main effect of critical item difficulty with no interaction would be observed on hit and false alarm rates of critical items (see Figure 2). More specifically, the main effect of critical item difficulty will show higher hit and false alarm rates of easy probes compared to that of difficult probes, independent of the list difficulty. For the noncritical items, a main effect of list difficulty with higher hit and false alarm rates for easy noncritical items would be observed based on the fluency attribution account, independent of the critical item difficulty (see Figure 3).

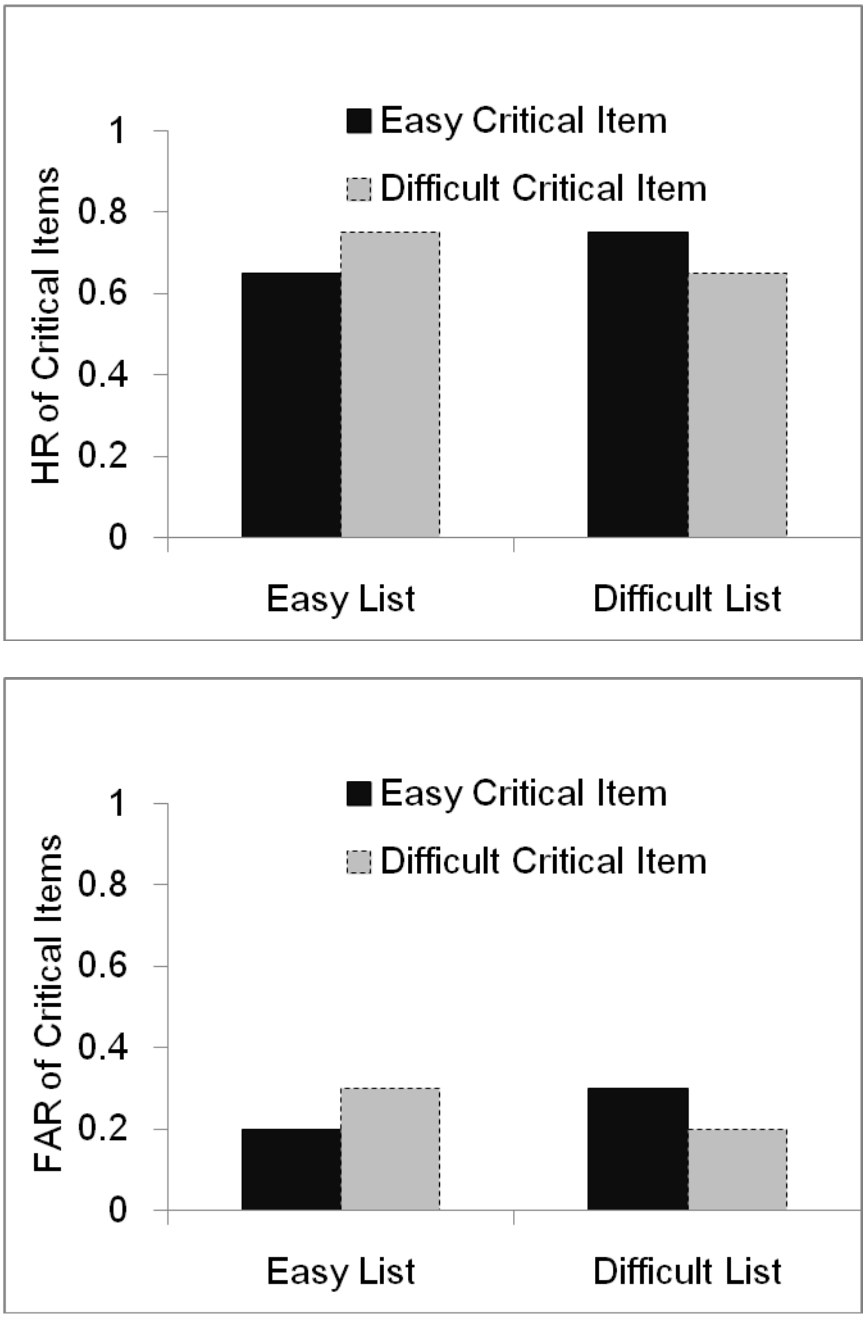

Figure 1. Predicted RM performance on the critical items by the discrepancy attribution account in Experiment 1a. HR = Hit rate, FAR = False alarm rate.

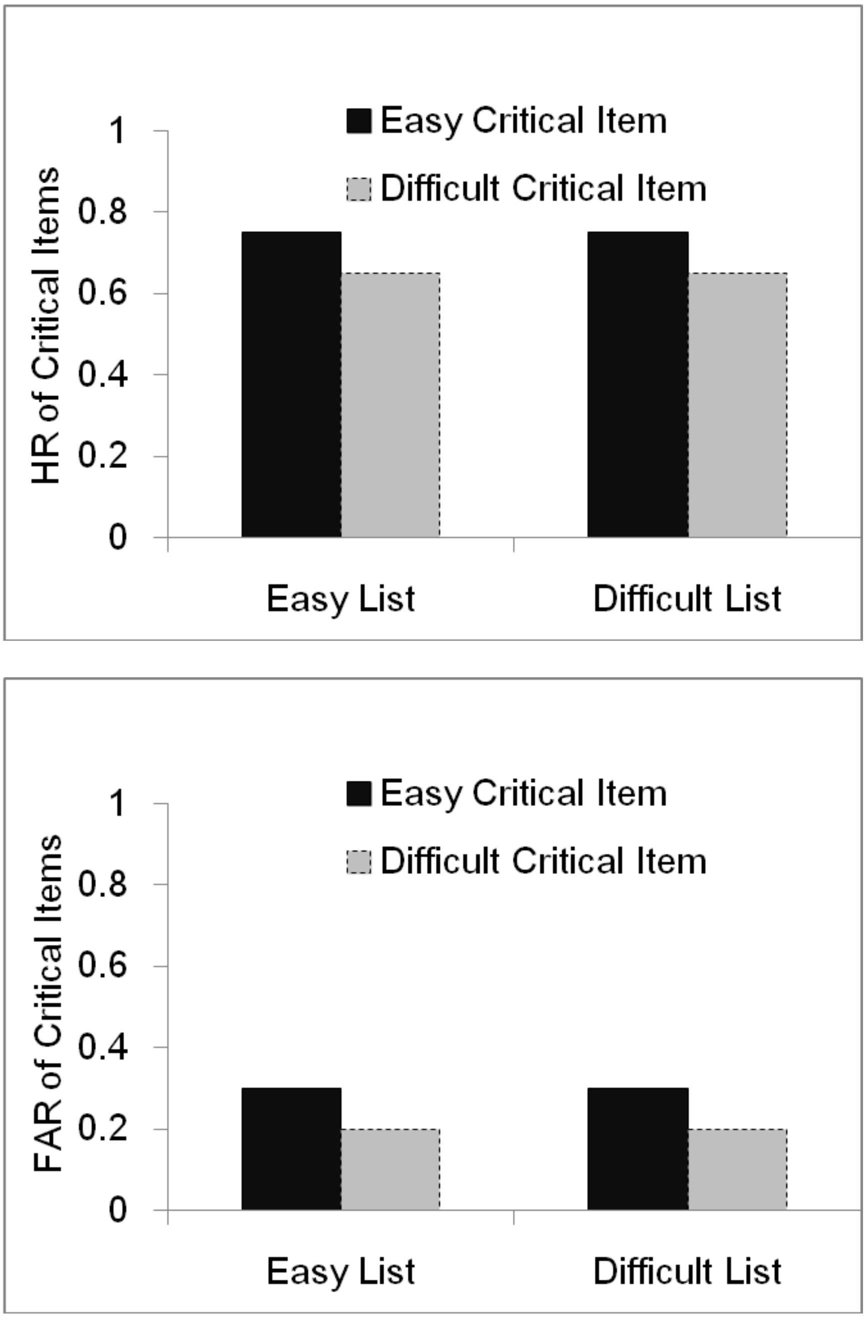

Figure 2. Predicted RM performance on the critical items by the fluency attribution account in Experiment 1a. HR = Hit rate, FAR = False alarm rate.

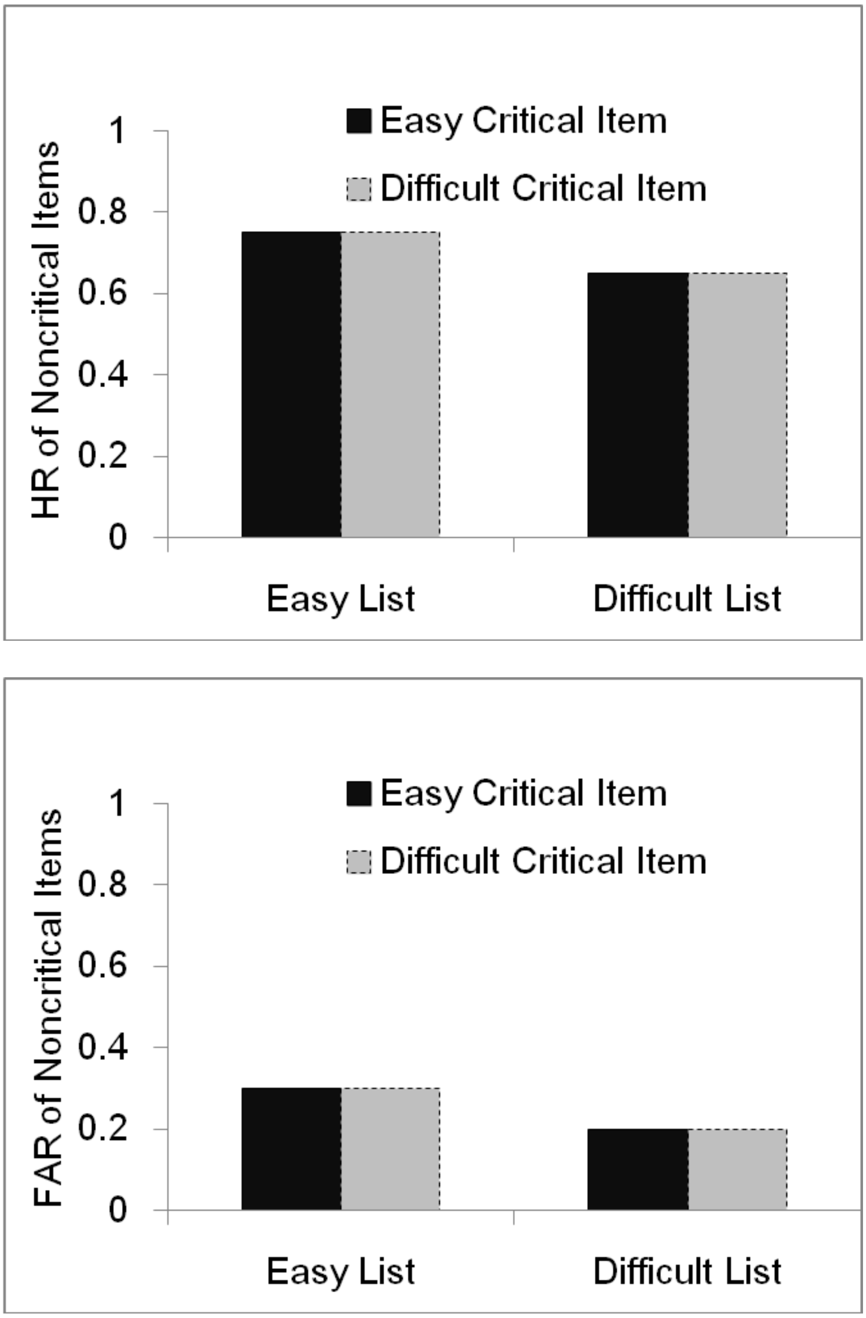

Figure 3. Predicted RM performance on the noncritical items by the fluency attribution account in Experiment 1a. HR = Hit rate, FAR = False alarm rate.

Although the discrepancy attribution account does not make a clear prediction for the recognition performance of noncritical items, one probable prediction is to find no main effect of noncritical item difficulty on both hit and false alarm rates of noncritical items (see

Figure 4). This prediction is postulated based on Whittlesea and Williams' (2001b) finding that showed no difference between recognition judgments on final words presented without a pause in the predictive versus consistent sentence stem. Even though final words in the predictive sentence stem were more predictive, hence, processed more fluently, the recognition judgments of final words in the predictive sentence stems did not differ from that of final words in the consistent sentence stems when no pause was inserted prior to the presentation of final word. Thus, it is also possible to find the same level of recognition judgments between the easy noncritical items and the difficult noncritical items in Experiment 1a.

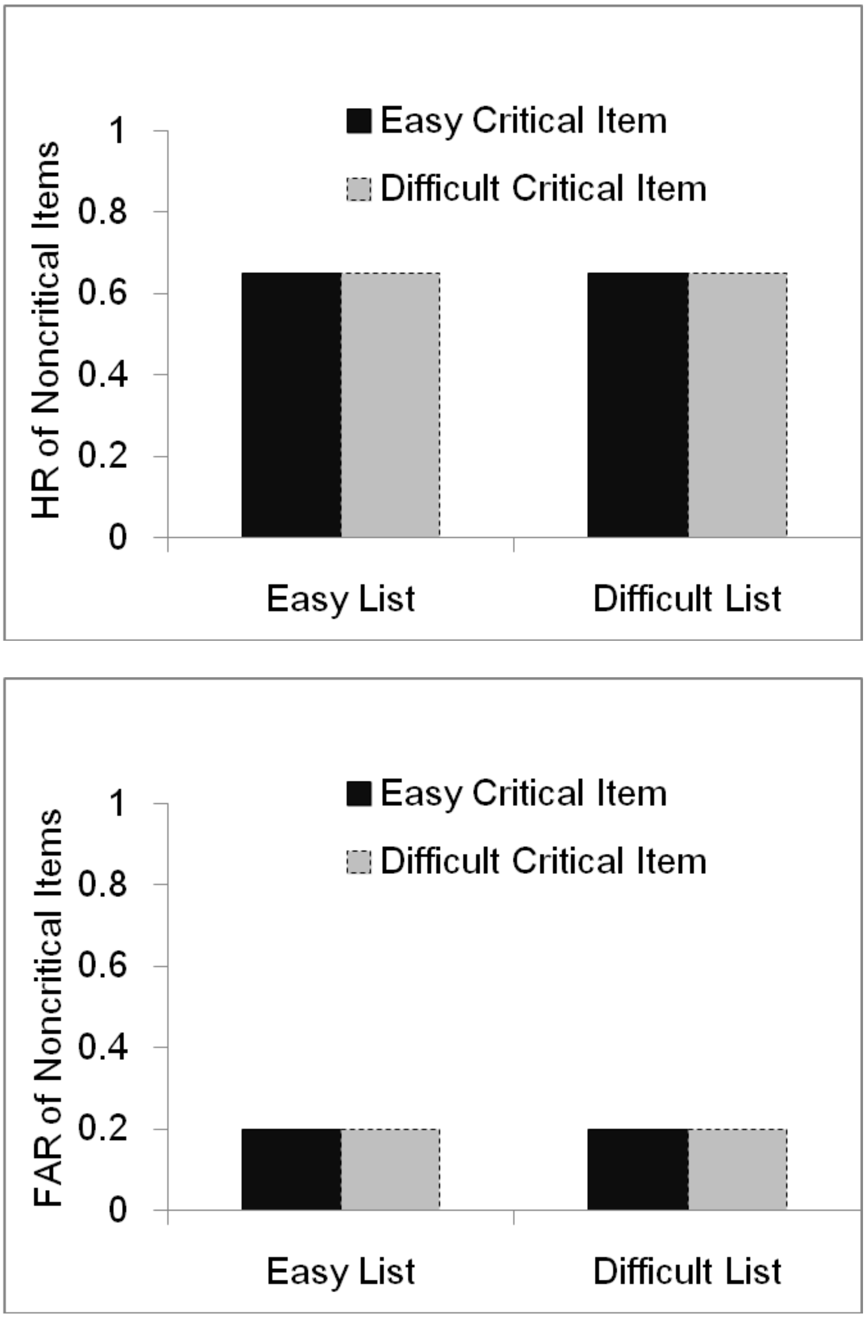

Figure 4. Predicted RM performance on the noncritical items by the discrepancy attribution account in Experiment 1a. HR = Hit rate, FAR = False alarm rate.

3.1 Method

3.1.1 Participants and design. The experiment was a 2 (Critical item difficulty; Easy/Difficult) X 2 (List difficulty; Easy/Difficult) X 2 (Study status; Studied/Nonstudied) X

2 (Lexicality; Word/Nonword) X 3 (Block type; First/Second/Third study-test block) mixed design. The critical item difficulty and the list difficulty were between subjects factors. Ninety-one participants were randomly assigned to one of four groups. There were 22 participants in the difficult critical items in the difficult list group and 22 participants in the easy critical items in the easy list group, with both groups reflecting the nondiscrepant condition. There were 19 participants in the easy critical items in the difficult list group and 28 participants in the difficult critical items in the easy list group, both groups reflecting the discrepant condition.

3.1.2 Materials. A set of 500 words was constructed (Scullin et al., 2010) as a stimulus set for the LDT. The set contained nouns, verbs, and adjectives with 4-8 letters and was divided into two subsets. Each subset had mean length of 6.5 and 6.47 and mean Log Hal Frequency (Balota et al., 2007) of 8.51 and 8.54. When one subset was used for word trials of the LDT, the other was used for nonword trials and this was counterbalanced. Nonword trials were constructed by changing the location of some letters from the base word or replacing some letters with a new letter(s). All nonwords were pronounceable. From these subsets, 6 lists of 36 words and 36 nonwords were constructed. With counterbalancing, half of these lists were used as study lists and the other half were used as nonstudied lists. Processing fluency defined as critical item difficulty and list (noncritical items) difficulty were manipulated by implementing two measures described below for both Experiments 1a and 1b.

Both easy and difficult test probes were presented in black font on a white background and a randomly generated string of colored (e.g., red) symbols (@#$%&*?8) was layered over them. The number of symbols matched that of the LDT probe and the color of layer was refreshed for each probe from four possible options (red, blue, brown, green). For difficult items, another layer of white symbols was inserted below the above-mentioned colored symbol layer (and above the black probe). Doing so led the difficult items to look as if white dots and colored symbols were masking some parts of the LDT probe in black font. Additionally, a non-pronounceable string of letters was used as a perceptually disrupting prime for difficult items with only the first and last letters of prime being the same as the LDT probe. For easy items, the white symbol layer was absent, thus, was less obscured visually, and the LDT probe itself was used as the identity prime (see Figure 5 for an example).

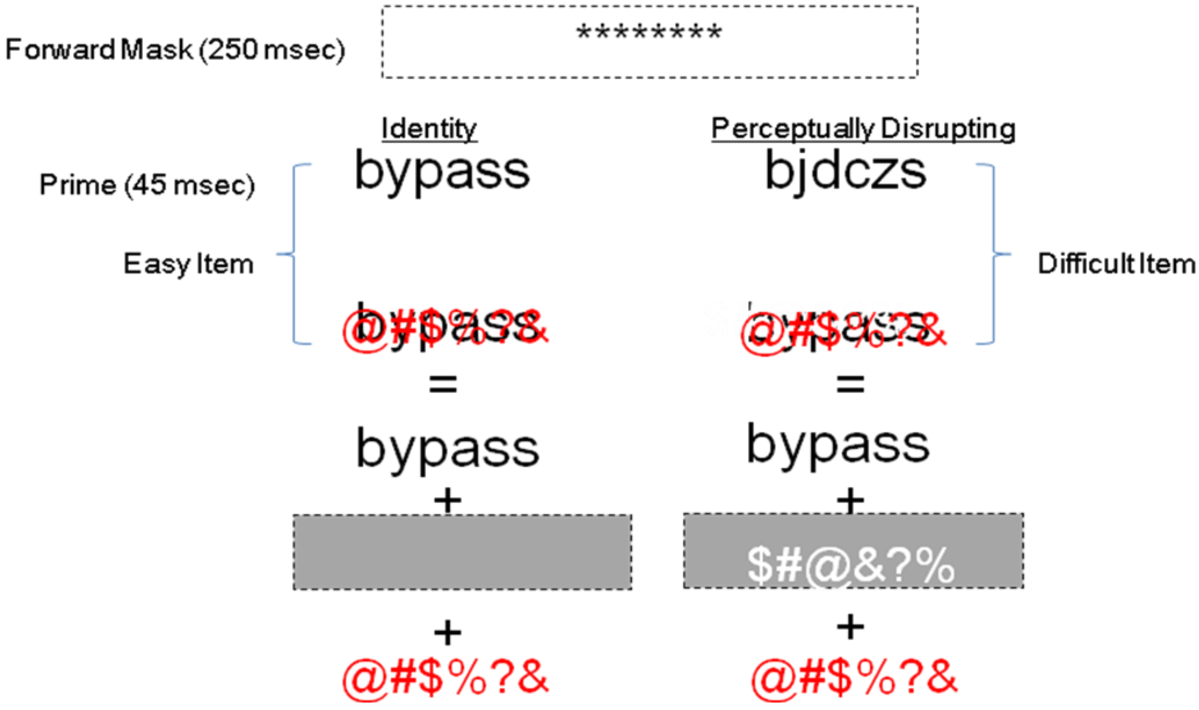

Figure 5. An example of stimulus presentation in Experiments 1a and 1b. The bottom of the figure depicts how each stimulus was constructed. Only the product of such construction, depicted above the equal sign, was presented to the participants. Dashed box indicates the common component of the stimulus presentation. The gray background is used to visualize the white layer of symbols. Easy item = a stimulus with no white layer of symbols + the identity prime. Difficult item = a stimulus with a white layer of symbols + the perceptually disrupting prime.

3.1.3 Procedure. Each participant performed three blocks of study and test phases with new study and distractor lists for each block. At the beginning of each block, participants were told to try to memorize the list of items to be presented. During the study phase of each block, a fixation signal of ******** was presented for 500 msec prior to each study item that was presented in black font in the white background. Each study item was presented for two seconds. A blank screen was presented for one second in between the study item and the subsequent fixation signal. Items from the study list(s) were presented randomly for each participant. Once participants completed studying the 72 items in each block (36 words and 36 nonwords), they engaged in a 90-second-long verbal distractor task (adapted and modified from the reading with distraction from Connelly, Hasher, & Zacks, 1991) to empty information from their working memory.

After the distractor task and prior to the first testing phase, the instruction for the LDT (that would be performed for each recognition probe) was provided. Participants were asked to decide whether a string of letters was a word or nonword by pressing the key labeled “y” for words and the “n” key for nonwords. The instruction was followed by a few practice trials with feedback on response accuracy and speed. For each practice LDT trial, a forward pattern mask (********) was presented for 250 msec followed by an appropriate prime of 45 msec. The practice trials were presented in the same manner as the actual test trials (e.g., a probe in black covered with the difficult perceptual mask that was overlaid with colored symbols with a disrupting prime as a difficult item) with no discrepancy. Half of the practice trials were words and the other half were pronounceable nonwords. All of the practice trials were items that were not studied or tested.

Upon completing the practice trials of the LDT, participants were told that the testing phase for recognition memory was about to begin. All of 72 studied items (36 words and 36 nonwords) within each block with the same number of distracters (36 words and 36 nonwords) were tested. The studied items and nonstudied items were intermixed pseudo-randomly within each testing phase so that no more than four trials of same study status or lexicality were presented in succession. During the testing phase, participants were first asked to judge if each item presented on the screen was a word or nonword with an appropriate key response. Then, they were asked if they remembered previously studying the item. The test list was constructed such that every 17th and 18th trials were the critical items and the mismatch/match of processing fluency difficulty of those trials to that of (16) preceding trials was the factor that induced (or did not induce) discrepancy. For all of the 17th and the 18th trials, in other words, critical items, study status was counterbalanced so that half of them were previously studied and the other were nonstudied to minimize a confounding effect of having two discrepant items sequentially. Furthermore, half of the critical items were words and the other half were nonwords. Upon completing all three study-test blocks, participants were asked a series of questions concerning if they were aware of the source of processing fluency, hence, discrepancy, and were able to attribute the discrepancy to its correct source (e.g., Whittlesea, 1993).

3.1.4 Data analysis. Data from the participants who commented on the white dot coverage of probes (suggestive of being aware of the perceptual mask manipulation) or the blinking of a screen (suggestive of being aware of the priming manipulation) during the post-experimental survey were excluded from the analysis. The reasoning for this exclusion was in line with the idea that being aware of the source of fluency or discrepancy precludes the misattribution of fluency or discrepancy to familiarity (e.g., Whittlesea, 1993). Of the 91 participants that were tested, 7 participants in the difficult critical items in the easy list group, 2 in the easy critical items in the easy list group and 2 in the difficult critical items in the difficult list group were excluded. This exclusion left 20 participants in the difficult critical items in the difficult list group, 20 in the easy critical items in the easy list group, 19 in the easy critical items in the difficult list group, and 21 in the difficult critical items in the easy list group.

All statistical tests were two-tailed with an alpha level of .05 unless noted otherwise for all experiments reported below. I report partial eta-squared for the Analysis of Variance (ANOVA) and Cohen's *d* as effect size measures where relevant. I also report Bayes information criterion (BIC) value as the posterior probability of the null hypothesis where relevant (Masson, 2011; Wagenmakers, 2007). Given the complexity of analyses with multiple Independent Variables (IVs), all analyses presented main effects first, followed by simple to higher-degree interactions. Also, for main effects, factors used to manipulate processing fluency were discussed first. For interactions, factors used to induce discrepancy were discussed first.

3.2 Results

3.2.1 LDT performance. First, as a manipulation check, I tested whether the processing fluency manipulation was successful by evaluating LDT performance (see Tables 1 and 2 for descriptive statistics).

Table 1.

Mean RTs of correct responses on LDT trials as a function of critical item difficulty,list difficulty, study status, lexicality, and item type in Experiment 1a

	Easy List		Difficult List	
Critical Item Difficulty	Easy	Difficult	Easy	Difficult
Critical Items				
Word				
Studied	1085 (194)	1306 (464)	1066 (160)	1257 (244)
Nonstudied	1138 (133)	1390 (781)	1129 (200)	1364 (481)
Nonword				
Studied	1329 (214)	1597 (556)	1351 (256)	1605 (597)
Nonstudied	1292 (251)	1601 (616)	1404 (272)	1701 (756)
Noncritical Items				
Word				
Studied	1031 (133)	1175 (447)	1262 (168)	1233 (185)
Nonstudied	1179 (162)	1199 (320)	1399 (230)	1374 (318)
Nonword				
Studied	1424 (157)	1451 (471)	1659 (346)	1702 (567)
Nonstudied	1352 (157)	1455 (429)	1645 (361)	1805 (828)

Note. RTs are in msec. Standard deviations are in parentheses.

Table 2.

Mean accuracy of LDT responses as a function of critical item difficulty, list difficulty, study status, lexicality, and item type in Experiment 1a

	Easy List		Difficult List	
Critical Item Difficulty	Easy	Difficult	Easy	Difficult
Critical Items				
Word				
Studied	.98 (.05)	.91 (.07)	.99 (.03)	.90 (.09)
Nonstudied	.96 (.06)	.86 (.14)	.95 (.06)	.88 (.12)
Nonword				
Studied	.89 (.11)	.83 (.12)	.81 (.14)	.85 (.16)
Nonstudied	.94 (.06)	.92 (.12)	.94 (.13)	.91 (.10)
Noncritical Items				
Word				
Studied	.95 (.06)	.92 (.09)	.87 (.09)	.87 (.10)
Nonstudied	.91 (.07)	.88 (.10)	.81 (.11)	.81 (.12)
Nonword				
Studied	.92 (.08)	.86 (.12)	.85 (.13)	.89 (.11)
Nonstudied	.95 (.12)	.91 (.08)	.87 (.12)	.90 (.09)

Note. Standard deviations are in parentheses.

LDT RTs of noncritical items. Mean raw LDT RTs of noncritical items were entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables (the block type was collapsed across three blocks)[1]. There was a main effect of list difficulty, *F* (1, 76) = 9.13, *MSE* = 448760.01, *p* = .003, η_p^2 = .11, such that LDT responses were faster for the easy items (*M* = 1283 msec) than for the difficult items (*M* = 1510 msec), showing that the processing fluency manipulation was successful. There was a main effect of study status, *F* (1, 76) = 12.08, *MSE* = 22947.42, *p* = .001, η_p^2 = .14, such that LDT responses were faster for the studied items (*M* = 1367 msec) than for the nonstudied items (*M* = 1426 msec). There was a main effect of lexicality, *F* (1, 76) = 103.585, *MSE* = 83953.90, *p* < .001, η_p^2 = .577, such that LDT responses were faster for words (*M* = 1232 msec) than for nonwords (*M* = 1562 msec).

The interaction of critical item and list difficulty was not significant, *F* < 1. There was a marginally significant interaction of list difficulty by study status, *F* (1, 76) = 3.729, *MSE* = 22947.42, *p* = .057, η_p^2 = .047. Individual contrasts showed that for the difficult list, LDT responses were faster for the studied items (*M* = 1464 msec) than for the nonstudied items (*M* = 1556 msec), *F* (1, 76) = 9.74, *MSE* = 17818.35, *p* = .002, although for the easy list, LDT responses did not differ between studied (*M* = 1270 msec) and nonstudied items (*M* = 1296 msec), *F* < 1. There was a marginal interaction of list difficulty by lexicality, *F* (1, 76) = 2.924, *MSE* = 83953.90, *p* = .091, η_p^2 = .037. Individual contrasts showed that LDT responses were faster for the easy list than for the difficult list, whether for words (Ms = 1146

---

[1] Similar analyses were conducted separately for words and for nonwords, instead of having “lexicality” as a factor, throughout the experiment where appropriate. The results of those analyses were consistent with the results of analyses with “lexicality” as a factor.

and 1317 msec, respectively) or for nonwords (*Ms* = 1421 and 1702 msec, respectively), *Fs* > 24.28. There was an interaction of study status by lexicality, *F* (1, 76) = 12.83, *MSE* = 17818.35, *p* = .001, η_p^2 = .144. Individual contrasts showed that LDT responses were faster for the studied words (*M* = 1175 msec) than for the nonstudied words (M= 1288 msec), *F* (1, 76) = 28.67, *MSE* = 171818.35, *p* < .001, although LDT responses for nonwords did not differ as a function of study status (Ms = 1559 and 1564 msec for the studied and nonstudied nonwords, respectively), *F* < 1.

There was a three-way interaction of critical item difficulty by study status by lexicality on LDT RTs of noncritical items, *F* (1, 76) = 6.87, *MSE* = 17818.35, *p* = .011, η_p^2 = .083. Individual contrasts showed that the LDT RTs of noncritical items were faster if they were previously studied than nonstudied, *Fs* > 3.36, except for the noncritical items that were nonwords in the conditions with the difficult critical items (see Figure 6). LDT RTs of noncritical items that were nonwords in the conditions with the difficult critical items did not differ whether they were previously studied (*M* = 1541 msec) or not (*M* = 1499 msec), *F* < 1.94. . No other main effects or interactions were significant, *Fs* < 1.6.

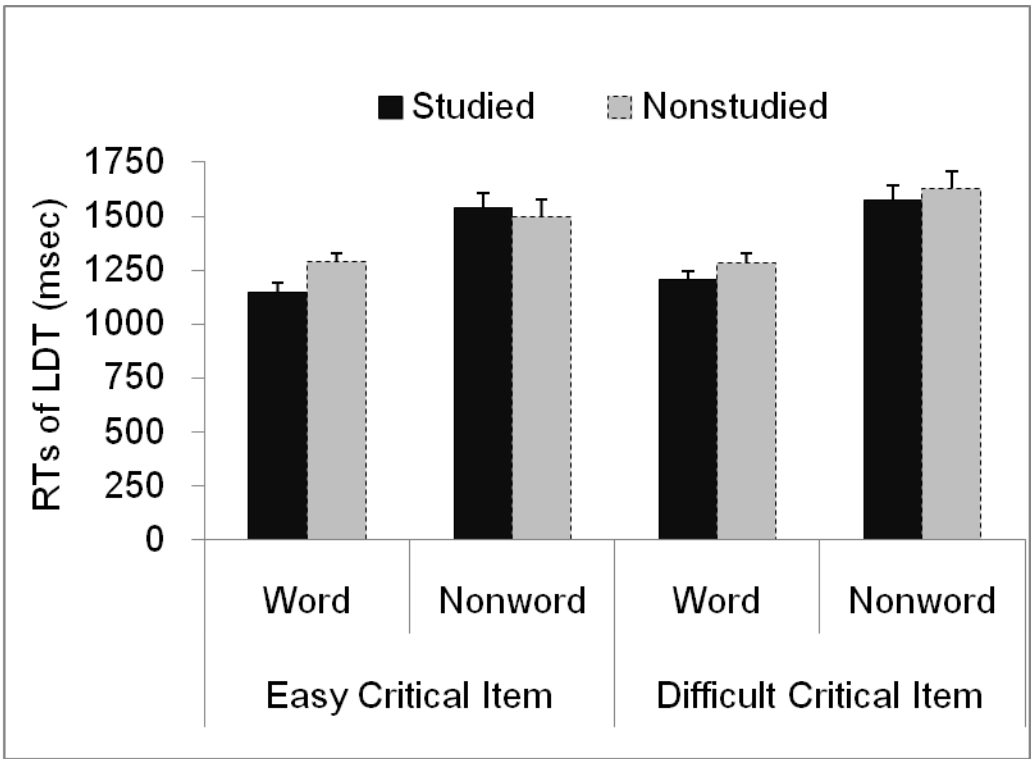

Figure 6. RTs of LDT of noncritical items as a function of critical item difficulty, study status, and lexicality in Experiment 1a. Error bars denote standard errors.

LDT accuracy of noncritical items. Mean LDT accuracy of noncritical items was also entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a main effect of list difficulty, F (1, 76) = 8.717, MSE = .028, p = .004, η_p^2 = .103, such that LDT accuracy was higher for the easy items (M = .91) than for the difficult items (M = .86), showing that the processing fluency manipulation was successful. There was a main effect of study status, F (1, 76) = 4.61, MSE = .001, p = .035, η_p^2 = .057, such that LDT accuracy was higher for the studied items (M = .89) than for the nonstudied items (M = .88).

The interaction of critical item difficulty by list difficulty was not significant, F (1, 76) = 2.28, MSE = .028, p = .135. There was an interaction of list difficulty by study status, F

(1, 76) = 6.79, *MSE* =.001, *p* = .011, η_p^2 = .082. Individual contrasts showed that LDT accuracy was higher for the studied items (*M* = .87) than for the nonstudied items in the difficult list (*M* = .85), *F* (1, 76) = 4.61, *MSE* = .001, *p* = .005, although study status did not affect the LDT accuracy of the easy list (*Ms* = .91), *F* < 1. There was an interaction of list difficulty by lexicality, *F* (1, 76) = 5.71, *MSE* =.007, *p* = .019, η_p^2 = .07. Individual contrasts showed that LDT accuracy was higher for nonwords (*M* = .88) than for words (*M* = .84) when they were difficult items, *F* (1, 76) = 64.0, *MSE* = .001, *p* < .001, and when they were easy items (*Ms* = .92 and .91 for nonwords and words, respectively), *F* (1, 76) = 4.0, *MSE* = .001, *p* = .05. There was an interaction of study status by lexicality, *F* (1, 76) = 80.46, *MSE* =.001, *p* < .001, η_p^2 = .514. Individual contrasts showed that LDT accuracy was higher for studied words (*M* = .90) than for nonstudied words (*M* = .86), *F* (1, 76) = 64.0, *MSE* = .001, *p* < .001, and for nonstudied nonwords (*M* = .91) than for studied nonwords (.88), *F* (1, 76) = 36.0, *MSE* = .001, *p* < .001. No other main effects or interactions were significant, *Fs < 1.3*.

LDT RTs of critical items. Next, I analyzed the LDT performance of critical items for the completeness of the analyses. However, caution is needed in interpreting these data, especially that of RTs, considering the limited number (maximum of 12) of observations per cell. Mean raw LDT RTs of critical items were entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a main effect of critical item difficulty, *F* (1, 76) = 9.33, *MSE* = 550430.64, *p* = .003, η_p^2 = .109, such that LDT RTs were faster for the easy items (*M* = 1224 msec) than for the difficult items (*M* = 1478 msec), showing that the processing fluency manipulation was successful. There was a marginally significant main effect of study status, F (1, 76) = 3.56, MSE = 62839.45, p = .063, η_p^2 = .045, such that LDT RTs were faster for the studied items (M = 1324 msec) than for the nonstudied items (M = 1377 msec). There was a main effect of lexicality, F (1, 76) = 57.55, MSE = 99784.16, p < .001, η_p^2 = .43, such that LDT RTs were faster for words (M = 1217 msec) than for nonwords (M = 1485 msec). No other main effects or interactions were significant, including the interaction of critical item difficulty by list difficulty, *Fs < 1.5*.

LDT accuracy of critical items. Mean LDT accuracy of critical items was also entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a main effect of critical item difficulty, F (1, 76) = 8.19, MSE = .023, p = .005, η_p^2 = .097, such that LDT accuracy was higher for the easy items (M = .93) than for the difficult items (M = .88), showing that the processing fluency manipulation was successful. There was a main effect of study status, F (1, 76) = 8.29, MSE = .006, p = .005, η_p^2 = .098, such that LDT accuracy was higher for the nonstudied items (m=.92) than for the studied items (m=.90). There was a main effect of lexicality, F (1, 76) = 12.76, MSE = .011, p = .001, η_p^2 = .144, such that LDT accuracy was higher for words (M = .93) than for nonwords (M = .89).

The interaction of critical item difficulty by list difficulty was not significant, F < 1. There was an interaction of critical item difficulty by lexicality, F (1, 76) = 7.76, MSE = .011, p = .007, η_p^2 = .093. Individual contrasts showed that although lexicality did not influence LDT accuracy of difficult critical items (Ms= .87 and .89 for words and nonwords, respectively), F < 1, LDT accuracy was higher for easy words (M = .97) than for easy nonwords (M = .90), F (1, 76) = 19.11, MSE = .005, p < .001. There was an interaction of study status by lexicality, *F* (1, 76) = 51.94, *MSE* =.005, *p* < .001, η_p^2 = .41. Individual contrasts found that LDT accuracy was higher for the studied words (*M* = .95) than for the studied nonwords (*M* = .85), *F* (1, 76) = 80.00, *MSE* = .005, *p* < .001, although LDT accuracy did not differ as a function of lexicality for nonstudied items (Ms = .91 and .93 for words and nonwords, respectively), *F* < 3.3. There was a four-way interaction of critical item difficulty by list difficulty by study status by lexicality, *F* (1, 76) = 5.61, *MSE* =.005, *p* = .02, η_p^2 = .069. No other main effects or interactions were significant, *Fs < 2.25*.

3.2.2 Recognition memory performance. Hit and false alarm rates were calculated separately for the critical items (every 17th and 18th trials) as well as for the noncritical items, in each block in all conditions. The block type was not collapsed across for the analysis of recognition memory because one can postulate that discrepancy might interact with the block type. (However, block type and lexicality were collapsed across in Table 3 for ease of presentation. See Table 3 for descriptive statistics and Figure 7 for the hit rates of the critical items and Figure 8 for the hit rates of the noncritical items from each group of participants.) One possibility is that recognition memory might decrease across blocks, due to increasing interference or fatigue, and consequently the reliance on familiarity in making recognition judgments might increase across blocks. This possibility would lead to a more pronounced effect of discrepancy on the later block than on the earlier block. Another possibility is that the strength of discrepancy might dissipate across blocks, due to increasing experience with discrepancy. This possibility would lead to a more pronounced effect of discrepancy on the earlier block than on the later block.

Table 3.

Mean proportion of 'Old' responses of recognition trials as a function of critical item difficulty, list difficulty, study status, and item type in Experiment 1a

	Easy List		Difficult List	
Critical Item difficulty	Easy	Difficult	Easy	Difficult
Critical Items				
Hit	.70 (.21)	.54 (.19)	.59 (.17)	.57 (.16)
False Alarm	.29 (.18)	.30 (.17)	.32 (.21)	.33 (.16)
Noncritical Items				
Hit	.66 (.16)	.60 (.17)	.52 (.13)	.59 (.16)
False Alarm	.29 (.17)	.33 (.14)	.29 (.12)	.32 (.13)

Note. Standard deviations are in parentheses.

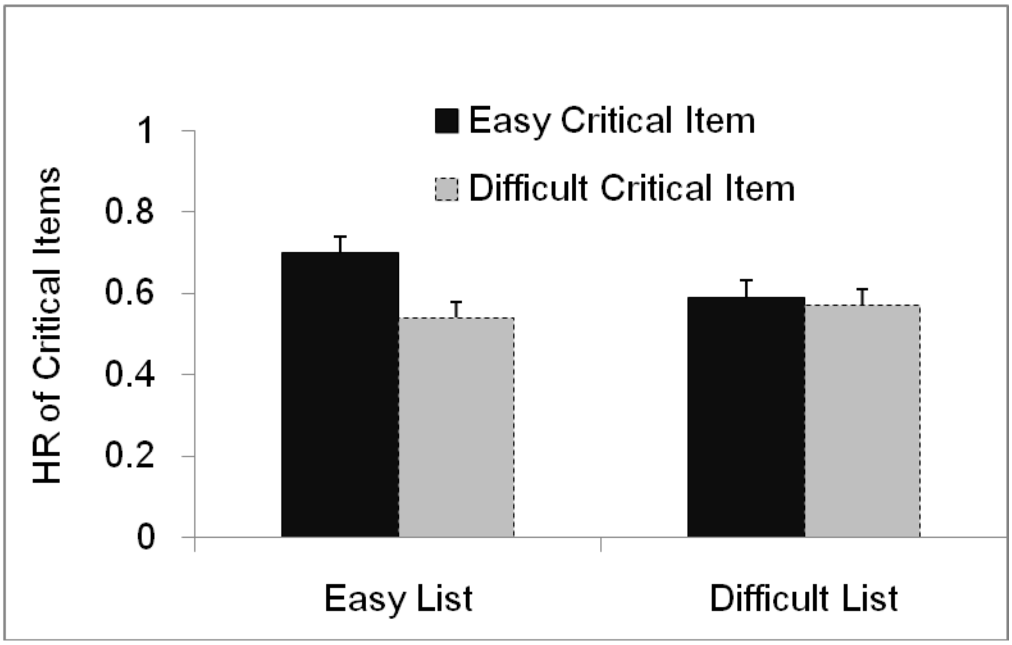

Figure 7. Actual RM performance on the critical items in Experiment 1a. HR = Hit rate, FAR = False alarm rate. Error bars denote standard errors.

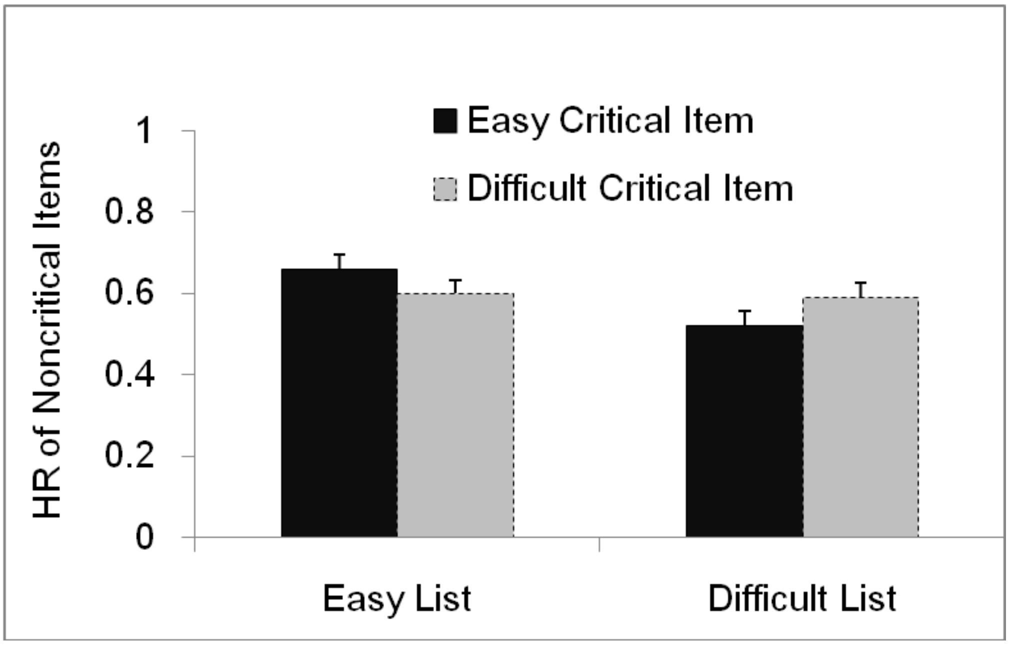

Figure 8. Actual RM performance on the noncritical items in Experiment 1a. HR = Hit rate. Error bars denote standard errors.

Hit rate of critical items. Proportion of correct “old” responses to previously studied critical items during the recognition test (critical item hit) was entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Lexicality: Word/Nonword) x 3 (Block: First/Second/Third) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a significant main effect of critical item difficulty, F (1, 76) = 4.83, MSE = .21, p = .031, η_p^2 = .06, such that the easy items were better remembered (M = .65) than the difficult items (M = .55). There was a marginal main effect of lexicality, F (1, 76) = 3.09, MSE = .082, p = .083, η_p^2 = .04, such that nonwords were better remembered (M = .62) than words (M=.58). There was a main effect of block type, F (2, 152) = 4.50, MSE = .059, p = .013, η_p^2 = .056. Individual contrasts found that previously studied items were remembered better in the first block (M = .65) than in the third block (M = .57), F (1, 152) = 4.27, MSE = .060, p = .04, and equally well in the third block and in the second block (M = .58), $F < 1$.

The interaction of critical item difficulty and list difficulty was not significant, F (1, 76) = 2.35, MSE = .21, p = .13. A Bayesian analysis showed weak support for no interaction, P_{BIC} (H_0|D) = .73, according to the guidelines set by Raftery (1995). Even if this interaction was significant, the pattern of data would not show support for the discrepancy attribution account because the highest hit rate (of this possible interaction) was observed from the easy critical items embedded in the easy list (see the far left bar from Figure 7). There was a marginal interaction of critical item difficulty by lexicality, F (1, 76) = 3.55, MSE = .082, p = .063, η_p^2 = .045. Individual contrasts showed that although hit rates of critical items did not differ as a function of critical item difficulty for words (*Ms* = .56 and .60 for the difficult and easy critical items, respectively, $F < 1$) or as a function of lexicality among the easy critical items (*Ms* = .60 and .70 for words and nonwords, respectively, F (1, 152) = 2.77, MSE = .06, p = .10), the nonword critical items were better recognized if they were easy (M = .69) than if they were difficult (M = .55), F (1, 76) = 6.53, MSE = .06, p = .01. No other main effects or interactions were significant, *Fs* < 2.63.

False alarm rate of critical items. Proportion of incorrect "old" responses to nonstudied critical items during the recognition test (critical item false alarm) were entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Lexicality: Word/Nonword) x 3 (Block: First/Second/Third) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a main effect of block, F (2, 152) = 3.10, MSE = .05, p = .048, η_p^2 = .04. Individual contrasts showed that the first block (M = .28) had a marginally lower false alarm rate than the third block (M = .34), F

(1, 152) = 2.88, *MSE* = .05, *p* = .09, but not compared to the second block (.32), *F* < 1. No other main effects or interactions were significant, including the interaction of critical item difficulty by list difficulty, *Fs < 1.82*. A Bayesian analysis showed positive support for no interaction, P_{BIC} (H_0|D) = .90.

Hit rate of noncritical items. Proportion of correct "old" responses to previously studied noncritical items during the recognition test (noncritical item hit) was entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Lexicality: Word/Nonword) x 3 (Block: First/Second/Third) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a marginal main effect of list difficulty, *F* (1, 76) = 3.89, *MSE* = .145, *p* = .052, η_p^2 = .05, such that the easy items were remembered better (*M* = .63) than the difficult items (*M* = .56). There was a significant main effect of block type, *F* (2, 152) = 29.92, *MSE* = .015, *p* < .001, η_p^2 = .28, such that previously studied items were better remembered in the first block (*M* = .65) compared to that in the second block (*M* =.58) and in the third block (*M* = .55), *Fs* (1, 152) > 21.78, *MSE* = .009, *ps* < .001.

There was a marginally significant interaction of critical item difficulty by list difficulty, *F* (1, 76) = 3.73, *MSE* = .145, *p* = .057, η_p^2 = .05. Individual contrasts found that for the difficult list (of noncritical items), memory for the noncritical items was higher in the condition where the difficult critical items were embedded (*M* = .59) than in the condition where the easy critical items were embedded (*M* = .52), *F* (1, 152) = 5.30, *MSE* = .009, *p* = .02, whereas for the easy list, memory for the noncritical items was higher in the condition where the easy critical items were embedded (*M* = .66) than in the condition where the difficult critical items were embedded (*M* = .60), *F* (1, 152) = 4.10, *MSE* = .009, *p* = .04.

Furthermore, in the conditions where the easy critical items were embedded, memory for the noncritical items was better for the easy list (*M* = .66) than for the difficult list (*M* = .52), *F* (1, 152) = 21.22, *MSE* = .009, *p* < .001. No other main effects or interactions were significant, *Fs < 2.23*.

False alarm rate of noncritical items. Proportion of incorrect "old" responses to previously nonstudied noncritical items during the recognition test (noncritical item false alarm) was entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Lexicality: Word/Nonword) x 3 (Block: First/Second/Third) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a marginal main effect of lexicality, *F* (1, 76) = 3.66, *MSE* = .027, *p* = .06, η_p^2 = .05, such that words were less falsely recognized (*M* = .29) than nonwords (*M* = .32). There was a main effect of block type, *F* (2, 152) = 9.96, *MSE* = .017, *p* < .001, η_p^2 = .12. Individual contrasts showed that the first block had a lower false alarm rate (*M* = .27), compared to the second block (*M* = .32) and the third block (*M* = .34); *Fs* (1, 152) > 14.29, *MSE* = .007, *ps* < .001.

There was no significant interaction of critical item difficulty by list difficulty, *F* < 1. There was a significant interaction of lexicality by block type, *F* (2, 152) = 6.23, *MSE* = .007, *p* = .003, η_p^2 = .08. Individual contrasts found that, in the first and second blocks, false alarm rates for words were lower (*Ms* = .25 and .29, respectively) than that for nonwords (Ms = .28 and .35), although false alarm rates for words (*M* = .33) and nonwords (*M* = .33) did not differ in the third block; *Fs* (1, 152) > 9.14, *MSE* = .007, *ps* < .003. No other main effects or interactions were significant, *Fs < 2.32*.

Given that the recognition judgments were preceded by the LDT judgment and the

LDT performance was not perfect, I also performed the above mentioned analyses taking the RM responses only from the correct LDT trials. The results were comparable to what was reported above, so I will not further discuss this conditional analysis.

Hit rates of critical and noncritical items in discrepant conditions. The above analyses tested the effect of the processing fluency manipulation separately for the critical items and the noncritical items and found no evidence of a discrepancy effect on recognition judgments of critical items. Another way to measure a possible discrepancy effect is to examine the effects of critical and noncritical item difficulty on recognition judgments of critical and noncritical items only from the discrepant conditions. When comparing the two groups of participants in discrepant conditions, theoretically competing predictions can be made. The fluency attribution account would predict a crossover interaction (see the two bars in the middle of upper panel in Figures 2 and 3). Given that the critical item difficulty and the noncritical item difficulty were mismatched for both groups of participants in the discrepant conditions, the conditional analysis of hit rates of critical and noncritical items from two discrepant conditions should lead to an interaction of item type by group type. That interaction would show higher hit rates of critical items from the group of participants who received the easy critical items in the difficult list than the group of participants who received the difficult critical items in the easy list and higher hit rates of noncritical items from the latter group than the former group. According to the discrepancy attribution account, however, no such interaction is predicted (see the two bars in the middle of upper panel in Figures 1 and 4). Instead, recognition judgments of noncritical items might not differ between the groups, leading to no main effect of group type. Furthermore, critical items that were discrepant should be judged more familiar than noncritical items, and thus, should lead to a main effect of item type with a higher familiarity for critical items compared to noncritical items.

To test these predictions, a conditional analysis was conducted with the dependent measure of proportion of correct “old” responses to previously studied critical and noncritical items only from participants in the discrepant conditions (those who received the easy critical items in the difficult list and the difficult critical items in the easy list). Proportion of correct “old” responses was entered into a 2 (Group type: the easy critical items in the difficult list group/ the difficult critical items in the easy list group) x 2 (Item type: Critical/Noncritical) mixed ANOVA with the group type as the between-participant variable (see the two bars in the middle of Figures 7 and 8). There was no significant main effect of group type or main effect of item type, *F*s < 1, the latter pattern disfavoring the discrepancy attribution account[2]. There was a significant interaction of group type by item type, F (1, 38) = 15.19, MSE = .005, $p < .001$, $\eta_p^2 = .29$. Individual contrasts showed those who received the easy critical items in the difficult list had a higher hit rate of critical items and a lower hit rate of noncritical items relative to those who received the difficult critical items in the easy list, *Fs* > 4.60. Therefore, these data patterns further corroborate the success of the processing fluency manipulation. Furthermore, these patterns are consistent with the higher hit rates of easy critical and noncritical items found in the more comprehensive ANOVAs above and provide support for the fluency attribution account.

RTs of correct recognition judgments of critical items and noncritical items. Next, I analyzed the RTs of correct recognition judgments. It is important to note that these data

---

[2] To inform the possibility that there was limited power to detect the discrepancy effect in Experiments 1a, 1b and 2, I report the power analyses for the three experiments in the general discussion.

should be interpreted with caution because of the reduced number of observations per cell for the critical and noncritical items resulting from the less-than-perfect mean accuracy of recognition performance. The block type was collapsed across three blocks. One participant was excluded from the analysis because no hit response was made on the word trials from this participant.

Mean raw RTs of correct recognition judgments of critical items were entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables (see Table 4 for descriptive statistics). There was a main effect of list difficulty, F (1, 75) = 5.21, MSE = 324408.86, p = .025, η_p^2 = .065, such that correct recognition judgments of critical items were faster in the conditions with the difficult list (M = 417 msec) than in the conditions with the easy list (M = 563 msec). There was a main effect of lexicality, F (1, 75) = 5.11, MSE = 96686.35, p = .027, η_p^2 = .06, such that correct recognition judgments were faster for words (M = 450 msec) than for nonwords (M = 529 msec). This main effect was qualified by a marginal interaction of lexicality by study status, F (1, 75) = 3.14, MSE = 76282.10, p = .080, η_p^2 = .04. Individual contrasts found that recognition judgments were slower for the nonstudied nonwords (M = 578 msec) than for the studied nonwords (M = 481 msec), F (1, 75) = 4.87, MSE = 76282.10, p = .03, although recognition judgments for the nonstudied words (M = 444 msec) were not slower than that for the studied words (M = 457 msec), $F < 1$. No other main effects or interactions were significant, $Fs < 1.63$.

Table 4.

Mean RTs of correct responses on recognition trials as a function of critical item difficulty, list difficulty, study status, lexicality, and item type in Experiment 1a

	Easy List		Difficult List	
Critical Item Difficulty	Easy	Difficult	Easy	Difficult
Critical Items				
Word				
Studied	589 (440)	532 (389)	361 (186)	353 (144)
Nonstudied	509 (271)	513 (342)	388 (233)	364 (182)
Nonword				
Studied	512 (308)	528 (348)	449 (242)	433 (209)
Nonstudied	574 (324)	755 (1035)	522 (328)	462 (188)
Noncritical Items				
Word				
Studied	502 (223)	447 (274)	393 (274)	391 (140)
Nonstudied	506 (249)	495 (319)	426 (189)	393 (163)
Nonword				
Studied Item	574 (254)	514 (278)	483 (328)	442 (177)
Nonstudied Item	707 (502)	572 (355)	517 (329)	487 (174)

Note. RTs are in msec. Standard deviations are in parentheses.

Mean raw RTs of correct recognition judgments of noncritical items were entered into a 2 (Critical item difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 2 (Study status: Studied/Nonstudied) x 2 (Lexicality: Word/Nonword) mixed ANOVA with the critical item difficulty and list difficulty as between-participant variables. There was a marginally significant main effect of list difficulty, F (1, 76) = 3.085, MSE = 251282.97, p = .083, η_p^2 = .039, such that correct recognition judgments of noncritical items were faster for the difficult list (M = 441 msec) than for the easy list (M = 540 msec). There was a main effect of lexicality, F (1, 76) = 24.38, MSE = 28141.54, p < .001, η_p^2 = .243, such that correct recognition judgments were faster for words (*M* = 444 msec) than for nonwords (*M* = 537 msec). There was a main effect of study status, *F* (1, 76) = 16.81, *MSE* = 9404.26, $p < .001$, $\eta_p^2 = .181$, such that correct recognition judgments were faster for the studied probes (*M* = 468 msec) than for the nonstudied probes (*M* = 513 msec). These main effects were qualified by a marginal interaction of study status by lexicality *F* (1, 76) = 2.85, *MSE* = 14486.46, $p = .095$, $\eta_p^2 = .036$. Individual contrasts found that recognition judgments were slower for the nonstudied nonwords (*M* = 571 msec) than for the studied nonwords (*M* = 503 msec), *F* (1, 76) = 12.40, *MSE* = 14486.46, $p < .001$, although recognition judgments for the nonstudied words (*M* = 455 msec) were not slower than that for the studied words (*M* = 434 msec), $F < 1.22$. No other main effects or interactions were significant, *Fs* < 2.27.

3.3 Discussion

Based on the faster RTs and higher accuracy of LDT for the easy items compared to the difficult items, it seems that the current paradigm successfully manipulated processing fluency with the perceptual mask modified from Whittlesea et al. (1990) and the masked prime. The main effects of processing fluency on hit rates of both critical and noncritical items (e.g., higher hit rates for the easy probes) also suggest that the manipulation was successful and provide further evidence for the fluency attribution account (Jacoby & Dallas, 1981). By contrast, the interactions of critical item difficulty and list difficulty on both hit and false alarm rates of critical items were not significant, thus there was no support for the discrepancy attribution account (Whittlesea & Williams, 1998, 2000, 2001a, 2001b). Based on the LDT data and the main effects of processing fluency on the hit rates of both critical and noncritical items, the failure to observe an effect of discrepancy on RM in this paradigm cannot be due to an unsuccessful processing fluency manipulation.

The main effect of block type on hit rates of both critical and noncritical items showed decreased performance across blocks, which seemed to reflect interference or fatigue over time. Decreasing recognition performance across blocks might cause a larger effect of discrepancy on recognition performance in the later block in which recognition was presumably more difficult compared to that in the earlier block in which recognition was relatively easier. This pattern is postulated based on findings that showed more reliance on discrepancy when the recognition task was more difficult (e.g., Goldinger & Hansen, 2006). Another possibility for a selective effect of discrepancy on a particular block would be a stronger effect of discrepancy in the earlier block than in the latter blocks, given that the early experience of discrepancy might be most potent. Both of these possibilities predict an interaction of critical item difficulty and list difficulty with block type on both hit and false alarm rates of critical items. However, the block type did not interact with any of the factors mentioned above for both hit and false alarm rates of critical items, providing no support for the aforementioned possibilities of discrepancy influencing recognition.

For noncritical items, I found a main effect of list difficulty that was qualified by an interaction of critical item difficulty and list difficulty on hit rates. The interaction suggested a possibility that hit rates were higher for the lists of noncritical items that contained the critical items of matching difficulty, whether easy or difficult, than the list of noncritical items that contained the critical items of mismatching difficulty. This pattern suggests that processing the critical items of mismatching difficulty might hinder recognition judgments of noncritical items. Conditional analysis of hit rates of both noncritical and critical items only in discrepant conditions showed that the easy noncritical items had a higher hit rate than the difficult noncritical items even when critical items of mismatching difficulty interfered with recognition judgments. This finding is inconsistent with the prediction of no difference between the easy and difficult noncritical items, postulated from the discrepancy attribution account. Rather, along with the higher hit rates of easy critical items, the higher hit rate of easy noncritical items of the conditional analysis further corroborates that the manipulation of processing fluency was successful at increasing "old" responses on recognition judgments for more fluently processed items (e.g., Jacoby & Whitehouse, 1989).

Although higher hit rates were found for both critical and noncritical items that were easy to process, false alarm rates for the easy items were not higher than that for the difficult items either for the critical items or the noncritical items. This finding is puzzling because Jacoby and Whitehouse (1989) claimed that processing fluency can influence familiarity judgments by showing higher rates of both hit and false alarm for more fluently processed items compared to less fluently processed items. Also, given that fluency can give rise to familiarity, if the effect of fluency was to be observed on RM, it would be more likely to be observed on the false alarm rate than on the hit rate, given that the former lacks recollection and hence is more prone to the influence of familiarity.

Given that the current project (1) used a string of white symbols as a perceptual mask instead of a layer of dots with varying density that Whittlesea et al. used (Whittlesea, Jacoby, & Girard, 1990), and (2) overlaid a string of colored symbols across test probes, it is possible that subjective experience of processing fluency in Experiment 1a might have differed from that in Whittlesea et al.'s experiments. It is possible that participants in Experiment 1a found the processing of test probes disfluent because of the unusual layer of colored symbols, regardless of the level of processing fluency manipulated. However, this possibility is unlikely, considering the RT measure of LDT that showed faster RTs for the easy items compared to that of the difficult items as well as the effect of processing fluency on hit rates. These RT measures have been used previously as a proxy for processing fluency by Whittlesea and colleagues (e.g., Whittlesea & Williams, 1998). Thus, the LDT data do not support the claim that the processing fluency manipulation of Experiment 1a did not translate into fluency, and hence, failed to increase false alarm rates.

Rather, it seems more parsimonious to conclude that the processing fluency of Experiment 1a was successfully manipulated and that fluency had an effect on hit rates but not on false alarm rates. Indeed, studies exist showing various factors can affect only recollection (e.g., speed of retrieval) or familiarity (e.g., perceptual matching of study and test) or both (e.g., study duration) (see Yonelinas, 2002 for a comprehensive review of these factors). Future research could investigate what aspect(s) of current paradigm allowed fluency to have an impact on the hit rate but not on the false alarm rate.

Another puzzling pattern in Experiment 1a was that the RTs of recognition judgments on both critical items and noncritical items in the conditions with the difficult list were faster than the RTs in the conditions with the easy list. Compared to the range of RTs reported in the literature (e.g., RTs of recognition judgments in the range of 1000 msec and RTs of LDT in the range of 600 ~ 800 msec, Duchek & Neely, 1989), the RTs of recognition judgments were relatively shorter (in the range of 400 ~ 700 msec) and the RTs of LDT were relatively longer (in the range of 1000 ~ 1800 msec) in Experiment 1a. Perhaps participants in Experiment 1a were able to think about the upcoming recognition judgments during the relatively longer period of LDT for the difficult probes, and hence were able to make faster recognition judgments for the difficult probes. Another possibility is that, given the relative disfluency of difficult probes, especially the difficult noncritical items, participants might not have exhibited enough effort into recognition judgments of the difficult probes which resulted in faster recognition judgments. This possibility aligns with the lower mean accuracy of recognition performance of difficult noncritical items.

The other possibility is that the relatively restricted numbers of observations used for the analyses of correct recognition judgment RTs and the between-participant design of list difficulty in Experiment 1a somehow generated an artifact. In support of this possibility, study status and lexicality (within-participant variables) showed significant effects in the direction consistent with the existing literature. The RTs of correct recognition judgments were faster for the studied probes than the nonstudied probes (Jou, Matus, Aldridge, Rogers, & Zimmerman, 2004) and faster for words than for nonwords (Rajaram & Neely, 1992). Nevertheless, neither of these possibilities can explain why the RTs of recognition judgments on critical items were slower in the conditions with the easy list than in the conditions with the difficult list. Unfortunately, the current project cannot distinguish the above mentioned possibilities.

Chapter 4: Experiment 1b

As mentioned previously, to perform a PM task, participants have to first recognize a stimulus as the PM cue at an appropriate time without being explicitly prompted. McDaniel et al. (2004) suggested that, discrepancy from a stimulus might be attributed as significance, rather than familiarity, of that stimulus under the PM context. They further suggested that this significance would initiate the search for the source of that significance and subsequently lead to the recognition of the PM cue. To test this suggestion, Experiment 1b used the same processing fluency paradigm Experiment 1a used and incorporated a PM task. The incorporation of a PM task with the paradigm that did not find a discrepancy effect in RM (Experiment 1a) was based on the premise that the utility of discrepancy might differ between PM and RM contexts. To perform a PM task, participants might not search for relevant information to guide their PM performance because they are absorbed in the ongoing activity. Furthermore, they might not access helpful information other than significance attributed from discrepancy to guide their PM performance. To perform a RM task, on the other hand, participants would actively search for information to guide their familiarity judgment and could access helpful information for that judgment, in addition to discrepancy, such as increased processing fluency due to the previous encounter. If the utility of discrepancy in the former condition is greater than that in the latter condition, it is possible to predict that the discrepancy-driven PM enhancement is more likely than the discrepancy-driven familiarity increase (see also Wänke, & Hansen, 2015, for the discussion of utility of perceiving changes in fluency across different cognitive tasks).

The intriguing prediction above has never been tested because no project has used the same paradigm to test the effects of discrepancy in PM and RM. Thus, Experiment 1b was the first study providing direct evidence regarding whether or not the same discrepancy manipulation influences PM but not RM. Also, Experiment 1b was the first experiment that used nonfocal PM cues in conjunction with a discrepancy manipulation. Below, I first briefly describe the definition of nonfocal PM cues and then the reasoning behind the use of nonfocal PM cues in Experiment 1b.

According to the multiprocess framework proposed by McDaniel and colleagues (Einstein et al., 2005; McDaniel & Einstein, 2000, 2007; McDaniel et al., 2004), a nonfocal PM cue refers to a PM cue whose features are not likely processed during the ongoing task (whereas a focal PM cue refers to a PM cue whose features are likely processed during the ongoing task). For example, if the ongoing task is to make a lexical decision for a string of letters, a PM cue of a word with the initial letter “o” will be considered as nonfocal (compared to a PM cue of a particular word “orange”). Because participants will focus on the lexical aspect of the presented word during that ongoing task, they do not need to attend to the initial letter of presented words (in the service of performing the ongoing task). Thus, the multiprocess framework suggests that for nonfocal PM cues, participants rely on strategic monitoring processes that are attention demanding, such as constantly checking if a stimulus for the ongoing task is the PM cue (e.g., McDaniel et al., 2004; Smith, 2003).

However, it is possible to conjecture that discrepancy might have an effect on nonfocal PM performance. Researchers suggest that strategic monitoring processes demand limited attentional resources (McDaniel & Einstein, 2007; Smith, 2003). If so, sustaining strategic monitoring processes for all possible PM cues across trials could be challenging. In support of this, studies using nonfocal PM cues often report PM performance that is lower than those using focal PM cues (e.g., Scullin et al., 2010, Experiment 4). If monitoring is not sustained throughout the PM task, thereby missing a number of PM cues, then discrepancy could assist the recognition of nonfocal PM cues that were missed by monitoring.

Experiment 1b used the processing fluency paradigm used in Experiment 1a and manipulated discrepancy by mismatching the expected and the actual difficulty level of processing fluency of PM cues during an ongoing task of LDT. The expectation (of difficulty of subsequent LDT trials) was built by having participants respond to a list of LDT trials with a particular processing difficulty, either easy or difficult. For example, those who responded to a list of easy LDT trials (that were paired with the identity prime and no perceptual mask described earlier in Experiment 1a) would presumably develop the expectation that the subsequent trials would also be easy to process. This expectation would be met by easy PM cues in the nondiscrepant condition and would be violated by difficult PM cues (that were paired with the disrupting prime and the difficult perceptual mask) in the discrepant condition. For the list of difficult LDT trials, the nondiscrepant condition presents the difficult PM cues and the discrepant condition presents the easy PM cues.

Often, the presence of a ceiling effect is one of the technical difficulties in studying any performance-enhancing factors in PM. To address this difficulty, Experiment 1b utilized multiple (six) trials of nonfocal PM cues. With the use of nonfocal PM cues and the difficulty with sustaining monitoring, PM performance in the nondiscrepant conditions is expected to be off the ceiling. If discrepancy plays a role in PM, participants should be more likely to make the PM response of pressing the “q” key while performing an LDT for the discrepant PM cues than for the nondiscrepant PM cues. In other words, both easy PM cues embedded in the difficult list and difficult PM cues embedded in the easy list would show higher PM performance than easy PM cues embedded in the easy list and difficult PM cues embedded in the difficult list (see Figure 9 for the predicted pattern of results).

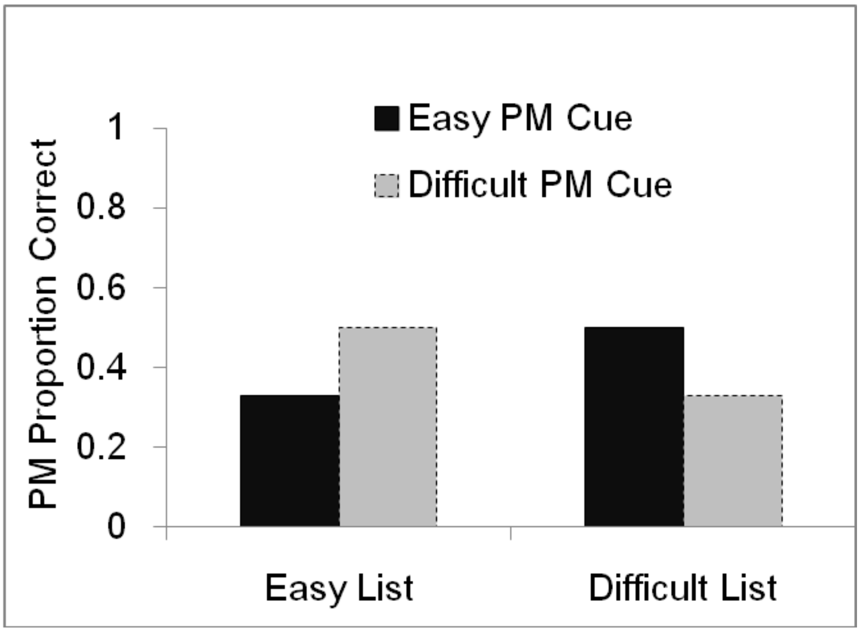

Figure 9. Predicted PM performance by the discrepancy attribution account in Experiment 1b.

To better inform the underlying mechanism(s) of discrepancy-enhanced PM performance, *monitoring costs* were measured along with the PM performance. Engagement in monitoring process is implicated by the relative slowing, *monitoring cost*, on performance of ongoing activity in the PM block (in which a PM task needs to be performed in addition to the ongoing task) compared to the performance of ongoing activity in the control block (in which only the ongoing task needs to be performed). Thus, in Experiment 1b, participants were asked to perform a control block of LDT as well as PM blocks. The overall monitoring costs were measured by comparing the mean RT of nontarget trials (that correspond to ongoing task trials that are not targets of the PM task) in the PM block to that in the control block.

4.1 Methods

4.1.1 Participants and design. The experiment was a 2 (PM cue difficulty;

Easy/Difficult) X 2 (List (of nontargets) difficulty; Easy/Difficult) X 4 (Block type; Control/First PM/Second PM/Third PM) mixed design. The block type was the only within-participant factor. Eighty-five participants were randomly assigned to one of four groups. There were 20 participants in the easy PM cue in the difficult list group and 22 in the difficult PM cue in the easy list group, both being discrepant conditions, and 22 in the difficult PM cue in the difficult list group and 21 in the easy PM cue in the easy list group, both being nondiscrepant conditions.

4.1.2 Materials. The same materials were used in Experiment 1b as in Experiment 1a with the following exception. Two sets of six PM cues were constructed with each set containing six words that have the same initial letter (husband, heritage, horizon, hero, hemp, herbal for words beginning with the letter "h"; mood, mundane, minority, monk, marker, magic for words beginning with the letter "m"). Mean length and Log Hal Frequency of words were 6 and 8.86 for the "h" set and 5.67 and 8.87 for the "m" set, respectively. Each participant received a set randomly drawn from the two possible options. Care was taken so that only the set of six PM cues had the specified initial letter throughout the experiment. All of PM cues were "yes" trials of LDT that were used as an ongoing task. Processing fluency defined as PM cue difficulty and list difficulty was manipulated by implementing the measures described under the Materials section of Experiment 1a.

4.1.3 Procedure. The instruction for LDT was provided and followed by a few practice trials with feedback on response accuracy and speed. Then, participants performed the control block of LDT that contained 50 trials, consist of 25 words and 25 nonwords. Upon completing the control block they were provided with the PM instruction asking them to press the "q" key if they ever see any words beginning with a specific letter (either "h" or “m”) during the experiment. After successfully repeating the PM instructions back to the experimenter, participants were asked to engage in a distractor task (adjusted for length from the materials used in Experiment 1a) for five minutes. Then, participants performed three PM blocks consecutively without being reminded of the PM instruction.

Each of three PM blocks had 164 trials. Two PM cues in each block were presented on the 107th and 158th trials. For discrepant PM cues, the PM cue difficulty mismatched the processing difficulty of preceding trials (also referred to as *list* difficulty). Thus, easy PM cues embedded in the difficult list (of nontargets) and difficult PM cues embedded in the easy list were considered discrepant PM cues. Easy PM cues embedded in the easy list and difficult PM cues embedded in the difficult list were considered nondiscrepant PM cues. Half of those trials in each block were the yes trial and the other half were the no trial for the LDT. Immediately after participants made their response for a given trial, the next trial appeared after a fixation signal. At the end of the third PM block, participants were surveyed with a series of questions regarding retrospective memory for the PM instructions and the detection of discrepancy and its attribution.

4.1.4 Data analysis. A total of 85 participants were tested and four of them were dropped as they commented on the white dot coverage of LDT probes or possible blinking of the screen, indicative of potentially aware of the processing fluency manipulations. Two participants from the difficult PM cue in the difficult list group, 1 from the easy PM cue in the difficult list group, and 1 from the difficult PM cue in the easy list group were excluded. This exclusion left 19 participants in the easy PM cue in the difficult list group, 20 in the difficult PM cue in the difficult list group, 21 in the easy PM cue in the easy list group, and 21 in the difficult PM cue in the easy list group.

4.2 Results

4.2.1 Prospective memory performance. All participants correctly recalled the PM instructions; thus, all participants were included in the analysis (cf. McDaniel, Shelton, Breneiser, Moynan, & Balota, 2011). The proportion of correct PM responses out of six trials was entered into a 2 (PM cue difficulty: Easy/Difficult) x 2 (List difficulty: Easy/Difficult) x 3 (Block type: 1st, 2nd, 3rd) mixed ANOVA with the block type as the within-participant variable. (See Table 5 and Figure 10 for descriptive statistics. Block type was collapsed across for the ease of presentation in the table and the figure.) The main effect of PM cue difficulty, F (1, 77) = .406, MSE = .309, p = .526, η_p^2 = .005, and the main effect of list difficulty, F (1, 77) = .655, MSE = .309, p = .421, η_p^2 = .008 were not significant. The interaction of PM cue difficulty by list difficulty was also not significant, F (1, 77) = .226, MSE = .309, p = .636, η_p^2 = .003. A Bayesian analysis showed positive support for no interaction on PM performance, P_{BIC} (H_0|D) = .90. This pattern of results held whether all PM responses were considered or only the PM responses on correct LDT trials were considered (Fs < 1).

Table 5.

Mean proportion of correct PM responses as a function of PM cue and list difficulty in Experiment 1b

	Easy List	Difficult List
Easy PM Cue	.24(.34)	.33(.33)
Difficult PM Cue	.25(.33)	.29(.36)

Note. Standard deviations are in parentheses.

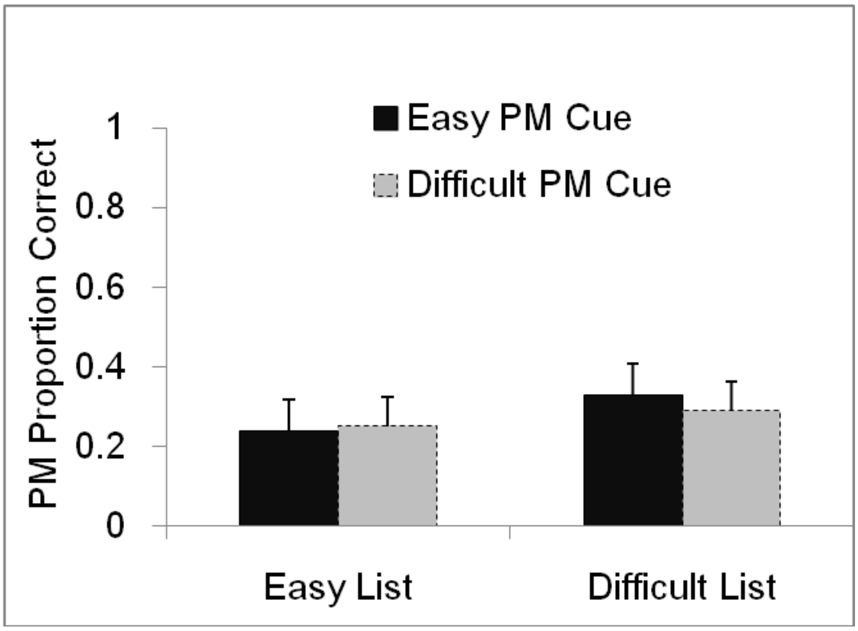

Figure 10. Actual PM performance in Experiment 1b. Error bars denote standard errors.

If discrepancy is induced by violation of expectation and the expectation can be subsequently adjusted across trials with experience during the experiment, the magnitude of discrepancy could decrease over PM trials. With this reasoning, I tested whether there was an effect of discrepancy on the first PM cue. First PM cue performance was better when it was embedded in the difficult list of LDT trials than in the easy list, $\chi^2(1, N = 81) = 4.08$, $p = .04$. PM performance for the first PM cue did not differ as a function of PM cue difficulty, $\chi^2(1, N = 81) = 1.42$, $p = .23$, or of discrepancy (collapsing across two discrepant and two nondiscrepant conditions across), $\chi^2(1, N = 81) = 1.09$, $p = .30$.

4.2.2 LDT performance. I next evaluated ongoing task performance to measure monitoring costs. Mean trimmed LDT RTs and mean LDT accuracy for nontargets were computed separately for the yes trials and the no trials for each block and analyzed appropriately. Given that the block type did not have any effect on PM performance in general, except that the list difficulty had an effect on the first PM cue performance, three PM blocks were collapsed across for the analysis of LDT RTs and accuracy.

LDT RTs. Mean trimmed RTs of nontargets (following Einstein et al's approach, 2005) was entered into a 2 (PM cue difficulty: Easy/Difficult) X 2 (List difficulty: Easy/Difficult) X 2 (Block type: Control/PM) X 2 (Lexicality: Word/Nonword) mixed ANOVA with the block type and lexicality as within-participant variables (see Table 6 for descriptive statistics). There was a significant main effect of list difficulty, F (1, 77) = 36.75, MSE = 193427.99, $p < .001$, $\eta_p^2 = .323$, such that LDT RTs were faster in the easy list (M = 834 msec) than in the difficult list (M = 1130 msec), suggesting that the processing fluency manipulation was successful. There was a significant main effect of lexicality, F (1, 77) = 113.85, MSE = 18954.55, $p < .001$, $\eta_p^2 = .597$, such that LDT RTs were faster for words (M = 900 msec) than for nonwords (M = 1064 msec). There was a significant main effect of block type, F (1, 77) = 4.85, MSE = 25930.45, $p = .031$, $\eta_p^2 = .059$, such that having a PM intention slowed down the RTs of LDT during the PM block (M = 1002 msec) than during the control block (M = 963 msec).

Table 6.

Mean RTs of correct LDT trials of nontargets as a function of PM cue difficulty, list difficulty, block type, and lexicality in Experiment 1b

	Easy List		Difficult List	
PM Cue Difficulty	Easy	Difficult	Easy	Difficult
Overall Monitoring Costs				
Word				
Control Block	776 (116)	750 (128)	1022 (188)	939 (148)
PM Block	782 (116)	794 (173)	1137 (289)	1004 (206)
Nonword				
Control Block	898 (175)	899 (197)	1284 (402)	1132 (271)
PM Block	864 (141)	908 (251)	1357 (500)	1169 (310)

Note. RTs are in msec. Standard deviations are in parentheses.

There was no significant interaction of PM cue difficulty by list difficulty, *F* (1, 77) = 2.25, *MSE* = 193428.0, *p* = .14. There was a marginally significant interaction of list difficulty by block type, *F* (1, 77) = 3.39, *MSE* = 25930.45, *p* = .069, η_p^2 = .042. Individual contrasts found that having a PM intention slowed down RTs of LDT in the difficult list (*Ms* = 1094 and 1167 msec for Control and PM blocks, respectively, *F* (1, 77) = 56.57, *MSE* = 3814.65, *p* < .001) but not in the easy list (Ms = 831 and 837 msec for Control and PM blocks, *F* < 1). There was an interaction of list difficulty by lexicality, *F* (1, 77) = 9.31, *MSE* = 18954.55, *p* = .003, η_p^2 = .11. Individual contrasts showed that LDT RTs were faster for words than nonwords in the easy list (Ms = 776 and 892 msec for words and nonwords, respectively) and in the difficult list (Ms = 1025 and 1236 msec for words and nonwords, respectively), *Fs* > 72.31. There was an interaction of lexicality by block type, *F* (1, 77) = 7.11, *MSE* = 3814.646, *p* = .009, η_p^2 = .085. Individual contrasts showed that monitoring costs were present for words (Ms = 872 and 929 msec for Control and PM blocks, respectively) and for nonwords (Ms = 1053 and 1074 msec for Control and PM blocks), *Fs* > 4.68. No other main effects or interactions were significant, *Fs* < 2.3.

LDT accuracy. Mean LDT accuracy of nontargets was entered into a (PM cue difficulty: Easy/Difficult) X 2 (List difficulty: Easy/Difficult) X 2 (Block type: Control/PM) X 2 (Lexicality: Word/Nonword) mixed ANOVA with the block type and lexicality as within-participant variables (see Table 7 for descriptive statistics). There was a significant main effect of list difficulty, *F* (1, 77) = 43.47, *MSE* = .017, *p* < .001, η_p^2 = .36, such that LDT accuracy was higher for the easy list (*M* = .93) than for the difficult list (*M* = .84), suggesting that the processing fluency manipulation was successful. There was a significant main effect of block type, *F* (1, 77) = 11.22, *MSE* = .003, *p* = .001, η_p^2 = .127, such that LDT accuracy was higher in the control block (M = .89) than in the PM block (M = .87).

Table 7.

Mean accuracy of LDT responses of nontargets as a as a function of PM cue difficulty, list difficulty, block type, and lexicality in Experiment 1b

	Easy List		Difficult List	
PM Cue Difficulty	Easy	Difficult	Easy	Difficult
Overall Monitoring Costs				
Word				
Control Block	.96 (.04)	.94 (.06)	.82 (.13)	.88 (.06)
PM Block	.94 (.04)	.93 (.04)	.75 (.14)	.82 (.07)
Nonword				
Control Block	.94 (.06)	.91 (.08)	.83 (.18)	.87 (.09)
PM Block	.93 (.02)	.90 (.11)	.87 (.07)	.84 (.11)

Note. Standard deviations are in parentheses.

There was a marginally significant interaction of PM cue difficulty by list difficulty, F (1, 77) = 3.10, MSE = .017, p = .08 (which was qualified by a three-way interaction described below). There was a marginal interaction of PM cue difficulty by lexicality, F (1, 77) = 3.44, MSE = .009, p = .067, η_p^2 = .083. Individual contrasts showed that in the easy PM cue conditions, LDT accuracy was lower for words (M = .87) than for nonwords (M = .90) at a trending level, F (1, 77) = 2.64, MSE = .005, p = .11, although in the difficult PM cue conditions, LDT accuracy were comparable between words (M = .89) and nonwords (M = .88), F < 1. There was also an interaction of list difficulty by lexicality, F (1, 77) = 6.96, MSE = .009, p = .01, η_p^2 = .083. Individual contrasts found that for the difficulty list, LDT accuracy was marginally lower for words (M = .82) than for nonwords (M = .85), F (1, 77) = 3.60, MSE = .005, p = .06, although for the easy list, lexicality did not matter for LDT accuracy (M = .94 for words and .92 for nonwords, F < 1).

There was an interaction of block type by lexicality, F (1, 77) = 7.56, MSE = .005, p = .007, η_p^2 = .089. Individual contrasts showed that having an PM intention of pressing the “q” key for specific words worsened the LDT accuracy of words (*Ms* = .90 and .86, for control and PM blocks, respectively), *F* (1, 77) = 12.96, *MSE* = .005, *p* < .001, although such intention did not influence the LDT accuracy for nonwords (*Ms* = .89 and .89 for control and PM blocks, *F* < 1). This interaction was qualified by a significant three-way interaction of list difficulty by block type by lexicality, *F* (1, 77) = 5.49, *MSE* = .005, *p* = .022, η_p^2 = .067. Individual contrasts showed that LDT accuracy was higher for nonwords (M= .85) than for words (*M* = .79) in the group in which easy PM cues were embedded in the difficult list, *F* (1, 77) = 6.76, *MSE* = .005, *p* = .011, whereas LDT accuracy for words and nonwords were comparable in the other three groups, *F*s < 1.64. No other main effects or interactions, except the three-way interaction described below, were significant, *Fs* < 2.6.

There was a marginally significant three-way interaction of PM cue difficulty by list difficulty by block type, *F* (1, 77) = 3.05, *MSE* = .003, *p* = .085, η_p^2 = .038. Planned comparisons of LDT accuracy of control versus PM blocks for individual groups were performed to test for monitoring costs. Participants in the difficult PM cue in the difficult list group had lower LDT accuracy in the PM block relative to the control block, *F* (1, 77) > 5.25, *MSE* = .005, *p* = .02, whereas LDT accuracy was the same between the control block and the PM block in the other three groups, *Fs* < 1.

4.3 Discussion

The RTs and accuracy data of LDT for nontargets in Experiment 1b replicated the results from Experiment 1a and further corroborated that the processing fluency manipulation of the paradigm was successful. PM performance in the nondiscrepant conditions was in the range (*Ms* = .24 ~ .29) that suggests monitoring was not sustained fully throughout the PM task. Nevertheless, I did not find any discrepancy-driven enhancement on PM performance.

In other words, PM performance on discrepant PM cues, whose processing difficulty mismatched the processing difficulty of the list they were embedded in, did not differ from PM performance on nondiscrepant PM cues, whose processing difficulty matched that of the list they were embedded in.

The multiprocess framework (McDaniel & Einstein, 2000, 2007) suggests that nonfocal PM cues are supported by monitoring processes rather than spontaneous retrieval. Considering that the discrepancy-induced significance is suggested as one possible mechanism for spontaneous retrieval, the use of nonfocal PM cues might have been detrimental to an attempt to find any discrepancy-enhancement on PM performance. Indeed, previous reports of discrepancy-enhancement on PM have used only focal PM cues (e.g., Lee & McDaniel, 2013; Thomas & McBride, 2015). However, as discussed previously, in the event where monitoring failed to detect nonfocal PM cues, as suggested by relatively low PM performance in the nondiscrepant conditions (*Ms* = .24 ~ .29), discrepancy could theoretically enhance nonfocal PM performance in the discrepant conditions. In the absence of discrepancy-enhancement on nonfocal PM performance in Experiment 1b, it is premature to determine whether discrepancy does or does not influence PM performance as McDaniel et al. (2004) have proposed. Future research comparing the effect of discrepancy on focal versus nonfocal PM cues may address this issue.

It was suggested that the utility of discrepancy might differ between PM and RM such that, even with the same paradigm, discrepancy might enhance performance in a PM task but does not increase familiarity in an RM task. However, no effect of discrepancy was observed on PM performance in Experiment 1b. Given that both Experiments 1a and 1b did not exhibit the effect of discrepancy, it is possible that my manipulation of discrepancy was somewhat incompatible to what Whittlesea and colleagues have used to induce discrepancy (e.g., Whittlesea & Williams, 1998). Based on the RTs and accuracy data of LDT from Experiments 1a and 1b, I was able to manipulate processing fluency, easy or difficult, with the perceptual processing paradigm. However, it is unclear whether presenting a series of items with a particular processing fluency and then presenting an item with the mismatching processing fluency was sufficient in eliciting discrepancy.

Indeed, the majority of Whittlesea's work that reported discrepancy effects has utilized expectations for processing fluency that were built pre-experimentally (e.g., Whittlesea & Williams, 1998, 2001a, 2001b). The majority of his work used three types of stimulus sets. The first consists of words and nonwords with orthographical regularity/irregularity, the second set consists of sentences ending with high/low constraint words, and the third set consists of semantically related/unrelated word pairs. Considering that most college students have very well-developed lexical and semantic knowledge, one can infer that from very early in those experiments, Whittlesea's participants had expectations of processing fluency of different stimuli and could have readily experienced violations of those expectations. In support of this possibility, Whittlesea and Williams (1998) showed that the higher false alarm rate for orthographically regular nonwords was found even when those nonwords were tested in a recognition memory test without the orthographically irregular nonwords. The reasoning was that the experience with orthographically irregular nonwords was unnecessary in finding the discrepancy effect for the orthographically regular nonwords.

There are some findings that can be interpreted under the discrepancy-attribution framework that have used sets of stimuli that are different from what Whittlesea and colleagues have used (e.g., picture of faces in Willems et al., 2007). Paradigms of those findings manipulated processing fluency of perceptual aspects of test probes, such as blurring/clearing of pictures (Willems et al., 2007) or decreasing/increasing color contrast between the test probe and its background (Hansen et al., 2008) and violated the expected processing fluency of some test probes to elicit discrepancy. The current project also attempted to elicit discrepancy by using a paradigm that manipulated perceptual processing fluency of test items and violating the expected processing fluency.

However, although the paradigms mentioned above were all successful at manipulating processing fluency, they might have differed with the ease of violating the expected processing fluency. Perhaps, the expectation for processing fluency was built and violated easier for clear/blurred pictures than stimuli used for the current project because detecting the difference in processing fluency of items in the former set was easier due to the greater pre-experimental experience with clear faces. In contrast, it is possible that the expectation for processing fluency was built rather slowly and violated less easily in the current project because detecting the difference in processing fluency was more challenging due to the lack of pre-experimental reference for how fluently items covered with colored symbols should be processed. Thus, although a small number of studies exist showing effects of discrepancy using paradigms other than those mentioned above (e.g., Lee & McDaniel, 2013), one explanation for the lack of discrepancy effect in Experiments 1a and 1b is that discrepancy is most pronounced in the violation of expectations for processing fluency that are built pre-experimentally. I explore this possibility in Experiment 2.

Experiment 1b focused on how discrepancy could enhance PM performance. Nevertheless, considering the suggestion that monitoring processes support nonfocal PM cues (Einstein et al., 2005; McDaniel & Einstein, 2000, 2007; McDaniel et al., 2004), I discuss the patterns of monitoring processes found in Experiment 1b and their implications to PM in general. Given that monitoring is resource-demanding (McDaniel et al., 2004; Smith, 2003), it is possible that participants performing a relatively easier ongoing task (e.g., LDT with a list of easy trials) might have more resources available to engage in resource-demanding, monitoring processes than those performing a relatively more difficult ongoing task (e.g., LDT with a list of difficult trials). Indeed, Lee and McDaniel (unpublished master's thesis) have shown that when participants are engaged in a less demanding ongoing task (e.g., category decisions with typical exemplars), they were more likely to engage in monitoring behaviors and to show higher focal PM performance compared to those engaged in a more demanding ongoing task (e.g., category decisions with atypical exemplars).

However, in Experiment 1b, opposite patterns of monitoring behaviors were observed. Based on the overall monitoring costs, participants in the difficult list groups seemed to engage in more monitoring behaviors than those in the easy list groups. One explanation for the greater monitoring costs in the difficult list groups is that the perceived difficulty of ongoing task might have encouraged participants in the difficult list groups to engage in more monitoring behaviors than those in the easy list groups. In support of this explanation, studies have found that participants' metacognition about the ongoing task difficulty and PM task difficulty can influence monitoring behaviors (e.g., Hicks, Marsh, & Cook, 2005). More comprehensive research is needed to explain what guides metacognitive decisions of monitoring behaviors in ongoing tasks with varying difficulty.

Although more overall monitoring costs were observed in the difficult list groups, PM performance was not greater in those groups than in the easy list groups with no overall monitoring costs. One possible explanation for the lack of PM enhancement in the difficult list groups with more overall monitoring costs is the efficacy of overall monitoring costs. Scullin et al. (2010) have suggested that overall monitoring costs alone might not be predictive of PM performance because they might be present even if proximal trials prior to a PM cue were not monitored. Indeed, when proximal monitoring costs were analyzed (following Scullin et al.'s approach), comparing the five trials preceding each PM cue to the corresponding trials in the control block, all four individual groups showed significant proximal monitoring costs. This pattern resonates with the comparable PM performance across discrepant and nondiscrepant conditions and is consistent with the idea that monitoring processes reflected in the proximal monitoring costs support nonfocal PM performance (e.g., Scullin et al.).

Chapter 5: Experiment 2

In Experiment 2, I sought to induce discrepancy by violating expectations for processing fluency that were built pre-experimentally. There were two factors I manipulated to induce differential processing fluency and, consequently, discrepancy in Experiment 2: word frequency and masked priming. Many studies have found that high and low frequency words differ with the ease of their processing, reflected in faster RTs and higher accuracy of high frequency words compared to low frequency words in an LDT (e.g., Balota & Chumbley, 1984; Rajarm & Neely, 1984). Furthermore, previous findings showed that people judge high frequency words as easier to process (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). These findings suggest that people are aware of the processing fluency advantage of high frequency words and the processing fluency disadvantage of low frequency words based on their pre-experimental experience. Thus, by using high and low frequency words as a stimuli set for Experiment 2, I can be more certain that my participants have pre-experimental expectations for processing fluency of the experimental stimuli as did participants in Whittlesea and Williams (1998) for orthographically regular nonwords and final words in sentences with the varying contextual certainty.

The other factor implemented in Experiment 2 was masked priming. As discussed previously, studies have shown that masked priming can influence the ease of processing (cf. Forster & Davis, 1984) and Experiments 1a and 1b showed supporting evidence for this. While Experiments 1a and 1b had two different types of primes, Experiment 2 had three types of primes. The first two types of primes were the identity prime of the recognition probe itself and the perceptually disrupting prime of a string of symbols (@#$%&?*8). The third type was a lexically disrupting prime, the word that is an orthographic neighbor of the recognition probe that has higher frequency than the probe word itself. Studies have shown that the use of an orthographic neighbor as a masked prime can hinder the processing of target stimulus (e.g., Segui & Granger, 1990). Use of two different types of disrupting primes could increase the chance of discrepancy induction whether the disruption of processing fluency in discrepancy induction is driven more by perceptual or lexical processes or both.

Another important advantage of using masked priming is that unlike unmasked priming, masked priming is less likely to exhibit the frequency attenuation effect. The frequency attenuation effect refers to differential facilitation of processing for high versus low frequency words, such as a greater benefit of priming for low frequency words (e.g., Rarajam & Neely, 1984). With masked priming, both high and low frequency words could show the effect of priming (although it is possible that the amount of facilitative priming is greater for low frequency words than high frequency words, see, Kinoshita, 1995).

By using high and low frequency words and manipulating the processing difficulty of those words via masked priming, I aimed to induce discrepancy by mismatching experienced processing fluency from that of expected processing fluency in a manner that closely resembled what Whittlesea and colleagues have done. In particular, Whittlesea and Williams (1998) have argued that higher false alarm rates for orthographically regular nonwords, compared to that of words or orthographically irregular nonwords, originate from the surprisingly more fluent processing of those nonwords afforded by orthographic regularity. They suggested that participants were not able to correctly attribute that surprisingly more fluent processing to orthographic regularity. I reasoned that the processing fluency differences of high and low frequency words and corresponding expectations would match those of words and orthographically regular nonwords. Furthermore, masked priming of facilitation and disruption would allow the manipulation of processing fluency of high and low frequency words without participants being aware of the source of the fluency variation. This overlap between Experiment 2 of the current project to that of Whittlesea and Williams led me to expect to find discrepancy effects on RM performance in Experiment 2.

Many studies have reported higher hit and lower false alarm rates for low frequency words than for high frequency words, which is called the word frequency mirror effect on recognition memory (e.g., Balota & Neely, 1980; Glanzer & Adams, 1985; Glanzer & Bowles, 1976; Jacoby & Dallas, 1981). The combination of the word frequency mirror effect in recognition memory and the processing fluency manipulation via masked priming allows competing predictions to be generated from the fluency attribution and the discrepancy attribution accounts. According to the fluency attribution account, fluency modulated by different types of primes could elevate the overall RM level and retain the word frequency mirror effect. More specifically, the fluency attribution account would predict a main effect of word frequency, showing the mirror effect described above on both hit and false alarm rates. Further, it would predict a main effect of prime type, such that test probes paired with the identity prime (producing more fluent processing) would have higher hit and false alarm rates compared to test probes paired with the disrupting primes. Lastly, the fluency attribution account would predict no significant interaction of word frequency and prime type on RM performance in the absence of the frequency attenuation effect (see Figure 11 for the predicted pattern of results).

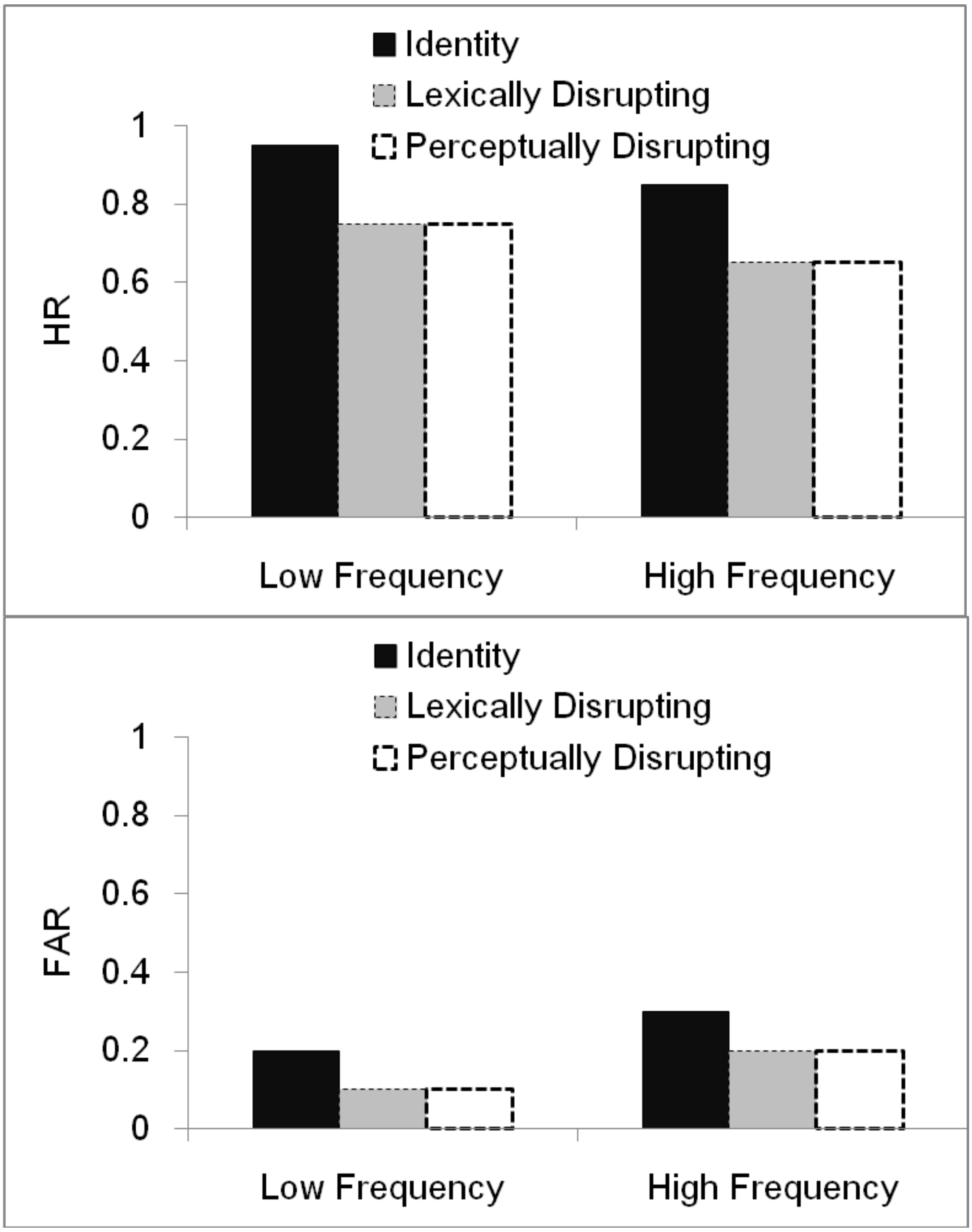

Figure 11. Predicted RM performance by the fluency attribution account in Experiment 2. HR = Hit rate, FAR = False alarm rate.

In the presence of the frequency attenuation effect, the identity prime might not be able to increase the processing fluency of high frequency words. If so, the identity prime might not increase hit and false alarm rates of high frequency words. Additionally, the disrupting primes might not be able to decrease the processing fluency of low frequency words, thereby, not decreasing hit and false alarm rates of low frequency words. In such cases, the differences on recognition performance among the different types of prime might decrease for both high and low frequency words. Nevertheless, the overall pattern of predicted results would not change (i.e., main effects of word frequency and prime type with no interaction of word frequency by prime type).

However, according to the discrepancy attribution account, a significant interaction of word frequency and prime type is predicted for both hit and false alarm rates. The reasoning is that participants might find the facilitation in processing fluency of low frequency words by the identity prime discrepant, because low frequency words are processed more fluently than anticipated. This discrepancy could give rise to familiarity, which would increase the hit and false alarm rates for low frequency words preceded by the identity primes relative to those preceded by the disrupting primes. In comparison, the disrupting primes could cause a discrepancy for high frequency words by violating the expected level of processing fluency for them. This discrepancy could then increase hit and false alarm rates of high frequency words preceded by disrupting primes relative to the high frequency words preceded by the identity primes (see Figure 12 for the predicted pattern of results). The presence or absence of the frequency attenuation effect, again, would not change the overall pattern of results predicted by the discrepancy attribution account.

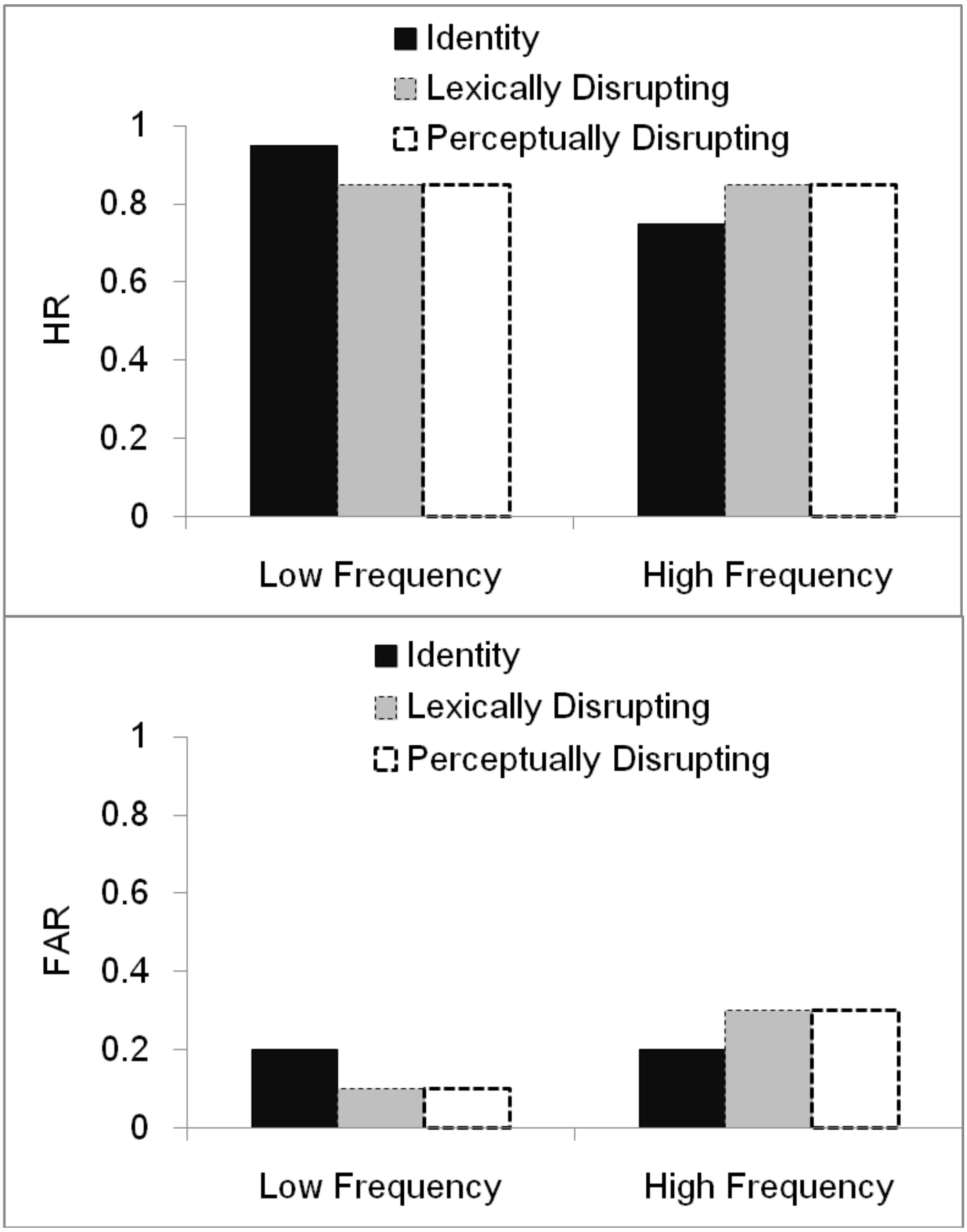

Figure 12. Predicted RM performance by the discrepancy attribution account in Experiment 2. HR = Hit rate, FAR = False alarm rate.

One well received explanation for the word frequency mirror effect on recognition memory by dual process theorists (e.g., Guttentag & Carroll, 1994; Joordens & Hockley, 2000; Reder et al., 2000) claims that higher hit rates of low frequency words compared to high frequency words are driven by differences in recollection. Furthermore, that explanation argues lower false alarm rates of low frequency words compared to high frequency words are driven by differences in familiarity. In support of this, studies have found more “remember” responses and fewer “know” responses to accompany hits of low frequency words compared to those of high frequency words (e.g., Gardiner & Java, 1990; Joordens & Hockley, 2000; Reder et al., 2000).

Both fluency and discrepancy are theorized to influence familiarity processes (e.g., Jacoby & Dallas, 1981; Whittlesea & Williams, 1998). Thus, if fluency were to influence RM performance as formulated above (i.e., a main effect of word frequency and a main effect of prime type on hits and false alarms without an interaction of word frequency by prime type), proportions of “remember” and “know” responses for low and high frequency words would replicate previous findings (e.g., more “remember” and fewer “know” responses to accompany hits of low frequency words compared to hits of high frequency words). Additionally, considering the possibility of fluency giving rise to familiarity, relatively more “know” responses are expected to accompany increased hits and false alarms of both high and low frequency words paired with the identity prime compared to those paired with the disrupting primes. However, if discrepancy were to influence RM performance, then a higher proportion of “know” responses should accompany hits and false alarms of low frequency words with the identity primes than those with the disrupting primes. Furthermore, a higher proportion of “know” responses should accompany hits and false alarms of high frequency words with the disrupting primes than those with the identity primes.

It is important to note that familiarity processes are more pronounced when the memory strength is weaker than when it is relatively stronger (e.g., Goldinger & Hansen, 2005). Thus, I aimed to have memory performance at a moderate level so that there was enough range for familiarity to be increased either by fluency or discrepancy. Below, I first present the pilot for Experiment 2 to show that the suggested manipulation is likely to be successful. I then describe Experiment 2.

5.1 Pilot

Seven participants were tested in a 2 (Word frequency; High/Low) X 3 (Prime type; Identity/Perceptually disrupting/Lexically disrupting) all within-participant design.

After consulting a number of articles on word frequency and recognition memory paradigms that reported their stimuli (e.g., Balota, Burgess, Cortese, & Adams, 2002), I constructed two sets of high and low frequency words, each set consisting of 48 low frequency words and 48 high frequency words. Adjectives, nouns, and verbs ranging between 3 to 7 letters were used. Mean length of words was 4.52 and 4.77 for high frequency and low frequency words, respectively. High frequency words had the mean Log Hal Frequency of 10.00 and low frequency words had the mean Log Hal Frequency of 7.08 (Balota et al., 2007). For each of these 192 words, I constructed three types of primes: the identity, the perceptually disrupting, and the lexically disrupting primes. The identity prime was the test probe itself. The perceptually disrupting prime was a string of symbols randomly constructed with @#$%&?*8 (e.g., “#$%&@” was used as the perceptually disrupting prime for the stimulus of “mayor”). The lexically disrupting prime was a word that was an orthographic neighbor of the test probe that had higher word frequency than the test probe itself (e.g., “major” was used as the lexically disrupting prime for the stimulus of “mayor”). Words used as the lexically disrupting prime for high frequency words had the mean Log Hal frequency of 11.42 and those used for low frequency words had the mean Log Hal frequency of 9.54.

One set of stimuli was used for word trials of the LDT. From the other set, 24 high frequency words and 24 low frequency words were selected and converted to nonwords that were pronounceable. The use of each set for word trials and the use of each prime type for each word were counterbalanced.

Participants were provided with instructions on how to perform the LDT. Then, they performed a block of practice trials with feedback on response accuracy and speed. Prior to the presentation of each LDT trial, a forward pattern mask of +*+*+*+* ("forward mask") was presented for 250 msec. This was followed by three presentations of an appropriate single prime for 15 msec each, alternated with two presentations of 15 msec-long midmask of @*#*$*@*. Total presentation time for the primes was 45 msec and the last prime was followed by the forward mask for 250 msec. The experimental block consisted of 96 words and 48 nonwords that were pseudo-randomly presented. No participant commented on the blinking of the screen during the post-experimental survey (see Figure 13 for an example stimulus presentation).

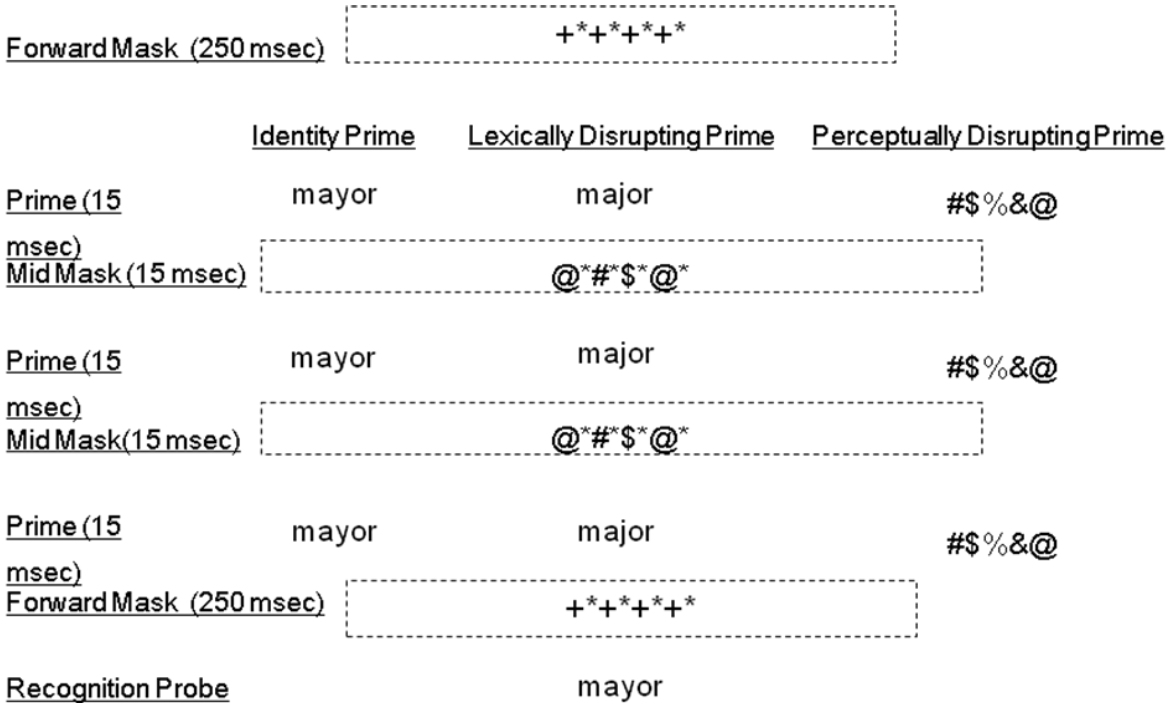

Figure 13. An example of stimulus presentation in Experiment 2. Dashed boxes indicate the common component of the stimulus presentation.

Mean trimmed LDT RTs of word trials were entered into a 2 (Word frequency: High/Low) X 3 (Prime type: Identity/Perceptually disrupting/Lexically disrupting) repeated measures ANOVA with all factors as within-participant variables (See Table 8 for descriptive statistics). There was a main effect of word frequency, F (1, 6) = 13.10, MSE = 1560.80, p = .011, η_p^2 = .69, such that LDT RTs were faster for high frequency words (M = 533 msec) than for low frequency words (M = 577 msec). There was a main effect of prime type, F (2, 12) = 8.31, MSE = 707.86, p = .005, η_p^2 = .581. Planned comparisons indicated that the identity prime led to faster LDT RTs (M = 531 msec) than either the lexically disrupting prime (M = 566 msec) or the perceptually disrupting prime (M = 568 msec), Fs (1, 12) > 5.23, MSE = 819.52, ps < .042. The interaction of word frequency by prime type was not significant, F < 1.

Table 8.

Mean RTs of correct LDT trials as a function of word frequency, prime type, and lexicality in the Pilot

Prime Type	Identity	Lexical	Perceptual
Word			
High Frequency	506 (55)	548 (54)	544 (87)
Low Frequency	556 (83)	583 (78)	591 (76)
Nonword			
High Frequency	657 (75)	654 (75)	674 (113)
Low Frequency	644 (72)	665 (94)	657 (110)

Note. “Lexical” refers to the lexically disrupting prime type and “Perceptual” refers to the perceptually disrupting prime type. RTs are in msec. Standard deviations are in parentheses.

Mean LDT accuracy of words trials was entered into a 2 (Word frequency: High/Low) X 3 (Prime type: Identity/Perceptually disrupting/Lexically disrupting) repeated measures ANOVA with all factors as within-participant variables (see Table 9 for descriptive statistics). The main effect of word frequency was marginally significant, F (1, 6) = 4.88, MSE = .003, p = .069, η_p^2 = .448, such that LDT accuracy was marginally higher for high frequency words (M = .98) than for low frequency words (M = .94). No other main effect or interaction of word frequency by prime type was significant, *Fs* < 1.75.

Table 9.

Mean accuracy of LDT responses as a function of word frequency, prime type, and lexicality in the Pilot

Prime Type	Identity	Lexical	Perceptual
Word			
High Frequency	.98 (.02)	.98 (.03)	.97 (.05)
Low Frequency	.95 (.04)	.95 (.03)	.92 (.03)
Nonword			
High Frequency	.91 (.09)	.84 (.17)	.84 (.19)
Low Frequency	.89 (.15)	.89 (.18)	.91 (.16)

Note. “Lexical” refers to the lexically disrupting prime type and “Perceptual” refers to the perceptually disrupting prime type. Standard deviations are in parentheses.

For the nonword trials of the LDT, there was no main effect of word frequency or prime type or interaction of word frequency by prime type on either mean RTs or accuracy of LDT, *Fs* < 2.08

The results of this pilot are consistent with the previous literature showing easier processing of high frequency words than low frequency words (e.g., Rajaram & Neely, 1992). Thus, these results suggest that experimentally valid high and low frequency words were selected. Furthermore, faster RTs of words preceded by the identity prime compared to RTs of words preceded by the lexically disrupting and perceptually disrupting primes suggest that the proposed priming manipulation of processing fluency was effective.

5.2 Methods

5.2.1 Participants and design. The experiment was a 2 (Word frequency; High/Low) X 3 (Prime type; Identity/Perceptually disrupting/Lexically disrupting) X 2 (Study status;

Studied/Nonstudied) all within-participant design. Fifty-three participants were tested.

5.2.2 Materials. Materials were the same as in the Pilot, with one exception. One set of 48 high frequency words and 48 low frequency words were used as study materials. The other set of 48 high frequency words and 48 low frequency words was used as nonstudied distracters during the test. Each set was equally used as study materials and distracters across participants.

5.2.3 Procedure. Participants first received the instruction for the recognition test and performed a block of practice trials. Then, for the experimental block, participants studied a list of 48 high frequency words and 48 low frequency words. Study words were presented in random order for each participant and each word was presented for two seconds. Upon the completion of the study phase, participants were asked to solve a list of math problems for 15 seconds as a distractor task.

After the study phase and the distractor task, participants were tested with all 96 studied words and 96 nonstudied words with matching characteristics. The test was pseudo-randomly ordered such that no more than four consecutive trials were of the same word frequency or prime type or study status. The recognition probe was presented in black font on a white background without any symbols covering over it (unlike Experiments 1a and 1b). Prior to the presentation of each recognition probe, a forward pattern mask of +*+*+*+* ("forward mask") were presented for 250 msec. This was followed by three presentations of an appropriate prime for 15 msec each, alternated with two presentations of 15 msec-long midmask of @*#*$*@*. Total presentation time for prime was 45 msec and the last prime was followed by the forward mask for 250 msec.

While performing the recognition test, participants were asked to make a remember/know response for the recognition probe they judged to be “old”. Explanations (adapted from Rajaram, 1993) were provided for when a remember response and when a know response was appropriate. At the end of the test block, participants were surveyed with a series of questions regarding the detection of discrepancy and its attribution.

5.2.4 Data analysis. Data from the participants who commented on the blinking of the screen and/or detection of prime during the post-experimental survey were excluded from the analysis. Of the 53 participants that were tested, 11 participants were excluded.

5.3 Results

5.3.1 Recognition memory performance. The hit and false alarm rates were calculated separately for trials for each word frequency (High/Low) and for each prime type (Identity/Perceptually disrupting/Lexically disrupting) (See Table 10 for descriptive statistics).

Table 10.

Mean proportion of “Old” responses on recognition trials as a function of word frequency, prime type, and study status in Experiment 2

	Prime Type	Identity	Lexical	Perceptual
Hit				
	High Frequency	.52 (.21)	.51 (.22)	.56 (.22)
	Low Frequency	.62 (.19)	.62 (.23)	.62 (.21)
False Alarm				
	High Frequency	.16 (.15)	.14 (.16)	.16 (.15)
	Low Frequency	.15 (.14)	.13 (.14)	.14 (.13)

Note. “Lexical” refers to the lexically disrupting prime type and “Perceptual” refers to the perceptually disrupting prime type. Standard deviations are in parentheses.

Hit rate. Proportion of correct “old” responses to previously studied items was entered into a 2 (Word frequency: High/Low) x 3 (Priming type: Identity/Perceptually disrupting/Lexically disrupting) mixed ANOVA with all factors as within-participant variables (see Figure 14). There was a main effect of word frequency, *F* (1, 41) = 38.43, *MSE* = .013, *p* = .031, η_p^2 = .06, such that low frequency words were better correctly recognized (.65) than high frequency words (*M* = .55). The main effect of priming type or interaction of word frequency by prime type were not significant, *Fs < 1.5*. A Bayesian analysis showed strong support for no interaction, P_{BIC} (H_0|D) = .95.

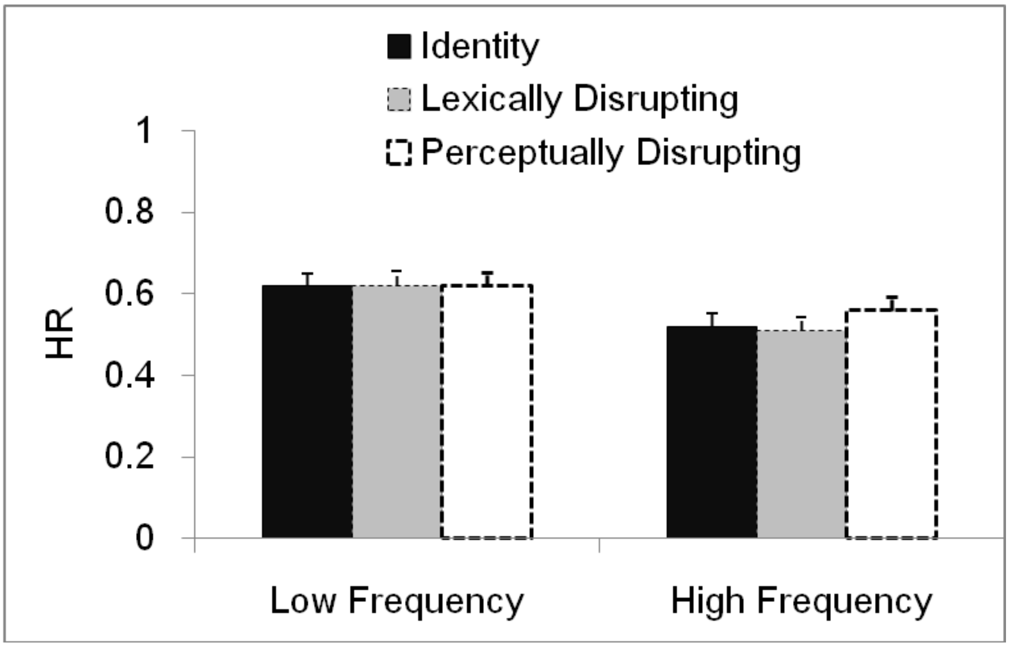

Figure 14. Actual RM performance in Experiment 2. HR = Hit rate. Error bars denote standard errors.

False alarm rate. Proportion of incorrect "old" responses to nonstudied items was entered into a 2 (Word frequency: High/Low) x 3 (Priming type: Identity/Perceptually disrupting/Lexically disrupting) mixed ANOVA with all factors as within-participant variables. There was no significant main effect or interaction, *Fs* < 1.6. A Bayesian analysis showed strong support for no interaction, P_{BIC} (H_0|D) = .98.

Remember/Know responses. Immediately after making the recognition memory judgment for each test word, participants were asked to make a “remember” or “know” response for the words they responded as having previously studied. Given the low level of false alarm rates and many participants not having a value for all possible cells, I will only report the proportion of “remember” and “know” responses for hits. Also, two participants did not have any hit for one of six cells. Thus, they were excluded from the analysis of remember/know responses and analysis of RTs of correct recognition judgments

Proportion of “remember” responses for “old” responses to previously studied items was entered into a 2 (Word frequency: High/Low) x 3 (Priming type: Identity/Perceptually disrupting/Lexically disrupting) mixed ANOVA with all factors as within-participant variables (See Table 11 for descriptive statistics). There was a main effect of word frequency, F (1, 39) = 7.31, MSE = .04, p = .01, η_p^2 = .16, such that the hits of low frequency words had a higher proportion of “remember” responses (M = .59) than the hits of high frequency words (M = .52). The main effect of prime type or interaction of word frequency by prime type was not significant, Fs < 1. Analysis of “know” response showed the reverse pattern such that there were more “know” responses for the high frequency words than that for the low frequency words.

Table 11.

Mean proportion of 'Remember' versus 'Know' responses to studied items correctly recognized during the recognition test as a function of word frequency and prime type in Experiment 2

	Prime Type	Identity	Lexical	Perceptual
Remember				
	High Frequency	.52 (.22)	.52 (.22)	.54 (.24)
	Low Frequency	.59 (.24)	.59 (.24)	.61 (.26)
Know				
	High Frequency	.48 (.22)	.48 (.22)	.46 (.24)
	Low Frequency	.41 (.24)	.41 (.24)	.39 (.26)

Note. "Conceptual" refers to the lexically disrupting prime type and "Perceptual" refers to the perceptually disrupting prime type. Standard deviations are in parentheses.

RTs of correct recognition judgments. Mean trimmed RTs of hit responses were entered into a 2 (Word Frequency: High/Low) X 3 (Prime type: Identity/Perceptually disrupting/Lexically disrupting) mixed ANOVA with all factors as within-participant variables (see Table 12 for descriptive statistics). There was a main effect of word frequency, *F* (1, 39) = 14.56, *MSE* = 28484.11, *p* < .001, η_p^2 = .27, such that hit responses for high frequency words were slower (*M* = 1191 msec) than for low frequency words (*M* = 1107 msec). There was a main effect of prime type, *F* (2, 78) = 3.54, *MSE* = 17995.80, *p* = .034, η_p^2 = .083, such that hit responses for words preceded by the identity prime were marginally faster (*M* = 1117 msec) than hit responses for words preceded by the lexically disrupting prime (*M* = 1172 msec), *F* (1, 78) = 3.39, *MSE* = 17839.71, *p* < .07, but not compared to hit responses for words preceded by the perceptually disrupting prime (*M* = 1158 msec). The interaction of word frequency by prime type was not significant, $F < 1$.

Table 12.

Mean RTs of correct responses on recognition trials as a function of word frequency, prime type, and study status in Experiment 2

	Prime Type	Identity	Lexical	Perceptual
Hit				
	High Frequency	1154 (313)	1215 (303)	1203 (238)
	Low Frequency	1080 (234)	1128 (215)	1114 (247)
Correct Rejection				
	High Frequency	1035 (276)	1044 (263)	1040 (289)
	Low Frequency	979 (225)	980 (209)	1022 (251)

Note. “Lexical” refers to the lexically disrupting prime type and “Perceptual” refers to the perceptually disrupting prime type. RTs are in msec. Standard deviations are in parentheses.

Mean trimmed RTs of correct rejection responses were entered into a 2 (Word Frequency: High/Low) X 3 (Prime type: Identity/Perceptually disrupting/Conceptually disrupting) mixed ANOVA with all factors as within-participant variables. There was a main effect of word frequency, F (1, 41) = 7.85, MSE = 17032.95, p = .008, η_p^2 = .16, such that correct rejections of high frequency words were slower (M = 1040 msec) than that of low frequency words (M = 994 msec). Main effect of prime type or the interaction of word frequency by prime type was not significant, *Fs* < 1.61.

5.4 Discussion

Using the materials and manipulation that showed different levels of processing fluency in the pilot, I found higher hit rates for low frequency words relative to that for high frequency words. I also found more “remember” and fewer “know” responses to accompany correctly remembered, previously studied low frequency words compared to that of high frequency words. Both of these findings are consistent with the existing literature that showed the hit advantage of low frequency words accompanied by more “remember” and fewer "know" responses (e.g., Gardiner & Java, 1990; Reder et al., 2000). Also replicating the existing literature are the faster RTs for hits of low frequency words than of high frequency words (e.g., Duchek & Neely, 1989). Despite replicating standard findings in the literature, I was not able to find discrepancy effects on RM. Below, I discuss the implications of these findings.

The pilot study verified that the priming manipulation produced an influence on the fluency of both low and high frequency words during an LDT: LDT RTs were faster for words with the identity prime than words with the perceptually and the lexically disrupting primes. Nevertheless, although there was a main effect of prime type on the RTs of hit responses, there was no main effect or any interaction of prime type on the mean accuracy of RM performance in Experiment 2. The possibility of no facilitative priming on high frequency words cannot easily account for this pattern of data. The present experiment was set up such that even if the identity prime does not enhance the processing fluency of high frequency words, the disrupting primes could reduce the processing fluency of high frequency words, to maximize the chance to observe a priming effect. The reasoning was similar for low frequency words such that even if the disrupting primes do not reduce the processing fluency of low frequency words, the identity prime could enhance the processing fluency of low frequency words. Even with such design features, however, there was no effect of priming on the accuracy of recognition judgments.

Given that Rajaram and Neely (1992) have reported that their use of masked priming of 50 msec increased hit and false alarm rates of both high and low frequency words in an intentional learning task, it is puzzling that hit and false alarm rates of low and high frequency words did not differ as a function of prime type in the present experiment. Given the possibility of the priming manipulation interacting with stimulus strength (e.g., Rajaram & Neely), which can be mapped to the study status of items, it is possible to predict a less strong effect of priming on studied words than nonstudied words and not find an effect of prime type on hit rates in Experiment 2. Even if this was the case, however, it would not interfere with finding an effect of prime type in the false alarm rates, which was absent in Experiment 2. Perhaps, a total of 45 msec of three 15-msec-long prime presentations (with two 15-msec-long midmask presentations interleaved in between) might not have been sufficient to show its effect on RM, although it successfully varied the fluency of responding in the LDT in the pilot study.

Even though the effect of multiple prime presentations of 15 msec in Experiment 2 might differ from the effect of a single prime presentation of 45 msec reported by Rajaram and Neely (1992), the lack of a priming effect on RM does not seem to be a sufficient explanation for the lack of the discrepancy effect on RM in Experiment 2. The main effect of priming type shown in the pilot, relative differences in processing fluency reflected in differential RTs of LDT, seems to closely resemble what Whittlesea and colleagues have presented as an explanation for higher false alarm rates for discrepant items. For example, Whittlesea and Williams (2001a) also found differential RTs in pronunciation of discrepant and nondiscrepant items. So, the lack of a priming effect on RM performance in Experiment 2 appears to be an insufficient explanation for the lack of discrepancy effect in Experiment 2.

Experiment 2 was conducted to explore one possibility for why a discrepancy effect was not found in Experiment 1a. One explanation for not finding the discrepancy effect on RM in Experiment 1a is the possible incompatibility of the expectations violated in Experiment 1a compared to most expectations used by Whittlesea and colleagues. The reasoning was that the expectation for processing fluency developed during the experiment by processing a list of easy or difficult items was not potent enough for its violation to cause discrepancy in Experiment 1a. On the other hand, the expectations Whittlesea and colleagues violated were built pre-experimentally over an extended period of one's life, such as expectations for how words and nonwords should be processed (1998) or how many words can sensibly fit in a particular sentence (2001b). To address this possible incompatibility, Experiment 2 attempted to induce discrepancy by violating the pre-experimental expectations for processing fluency of low and high frequency words with the use of different primes.

Nevertheless, even with the manipulations that could violate the pre-experimental expectations (as confirmed by the RT pattern on the LDT in the pilot), there was no effect of discrepancy on RM in Experiment 2. Perhaps, participants develop multiple expectations in processing items for a recognition task. Some studies have found that participants expect that high frequency words will be better remembered than low frequency words on a subsequent recognition test when asked to judge the memorability of items at study although this pattern reverses when the judgment was made at test (Benjamin, 2003; Guttentag & Carroll, 1998). Furthermore, McCabe and Balota (2007) reported that experience with high or low frequency words at study or test alone can influence the placement of decision criterion for a recognition judgment for medium frequency words. Their report suggests that people can adjust expectations for memorability of the same items in comparison to other items.

However, prior experience of a recognition task alone does not lead to the reversed pattern of memorability judgments for high and low frequency words, but instead leads to comparable levels of memorability judgments for high and low frequency words (Experiment 2 of Benjamin, 2003). Thus, even if the expectation of memorability of recognition probes was to play a role in Experiment 2, the combination of expectations for processing fluency and memorability of high and low frequency words should have been consistent with expectations for processing fluency alone. Unfortunately, it is unclear what the discrepancy attribution account would predict when multiple expectations might contradict one another and which expectation would take priority. Given this uncertainty of predictions from the discrepancy attribution account, it seems most parsimonious to conclude that, together with the null effect of discrepancy on RM in Experiment 1a, the results of Experiment 2 corroborate the claims that suggest discrepancy might not be as a general phenomenon as Whittlesea and colleagues have proposed (Cleary et al., 2007; Karpicke et al., 2008).

Although I found an advantage of low frequency words in hit rates accompanied by more “remember” responses, there was no effect of word frequency on false alarm rates, which seems surprising at first glance. Although most papers on word frequency and recognition memory report the usual word frequency mirror effect (higher hit and lower false alarm rates of low frequency words), several papers showed deviations from this pattern. Those papers manipulated factors that are known to differentially influence recollection and familiarity processes, such as speeded response deadline (e.g., Experiment 2 in Balota et al., 2002; Bridger, Bader, & Mecklinger, 2014), which eliminates the hit advantage but retains the false alarm advantage of low frequency words.

However, a closer inspection of the existing literature shows that a number of studies report no false alarm advantage for low frequency (relative to high-frequency) words in the presence of a hit advantage for low frequency words (Experiments 1 and 2, Coane, Balota, Dolan, & Jacoby, 2011; Experiment 1, Gardiner & Java, 1990; Experiment 1, Rajaram & Neely, 1992; Experiments 1a, 1b, and 2 of Pazzaglia, Staub, & Rotello, 2014). Both Rajaram and Neely and Pazzaglia et al. attributed their failure to find the false alarm advantage of low frequency words to methodological variations of their studies compared to other studies that have reported both hit and false alarm advantage of low frequency words. In light of the same pattern of data from Experiment 2, it is unlikely, however, that the lack of a false alarm advantage for low frequency words is due to random methodological variations in these studies. Perhaps, this pattern suggests a boundary condition of the word frequency mirror effect that can be further investigated.

Experiment 2 found that RTs of hit responses for the recognition probes preceded by the identity primes were faster than those preceded by the lexically disrupting primes. Although this pattern seems to suggest a successful priming manipulation, it is difficult to interpret its implication. Faster RTs of recognition judgments could be interpreted as indicative of stronger memory strength for recognition probes (e.g., Jou et al., 2004). Perhaps, the enhanced processing fluency of recognition probes by the identity prime were interpreted as representing stronger memory strength and led to faster RTs of recognition judgments. If this was the case, this would support the fluency attribution account which suggests that the fluent processing of recognition probes can increase the familiarity judgments for those items. However, the faster RTs of hit responses for the recognition probes preceded by the identity primes were not accompanied by higher hit responses. Also, because the RTs of recognition judgments measured the time it took participants both to perceive a recognition probe and to make a recognition judgment, it is difficult to distinguish whether the priming effect on the RTs of recognition judgments reflects the priming effect on the perceptual processes or on the recognition processes or both in Experiment 2.

Chapter 6: General Discussion

Some researchers have claimed that discrepancy could be attributed as familiarity and influence recognition memory (Whittlesea & Williams, 1998, 2000, 2001a, 2001b) and false memory (Whittlesea, 2002; Whittlesea et al., 2005). Other researchers have extended this claim and proposed that discrepancy could be attributed as significance and influence PM (Lee & McDaniel, 2013; Thomas & McBride, 2015). However, other studies have found evidence against discrepancy influencing recognition memory (Cleary et al., 2007) or false memory (Karpicke et al., 2008). Based on these opposing claims and data, the current project investigated the generality of the discrepancy effect on RM and PM using paradigms that manipulated actual processing fluency and expected processing fluency, either developed during (Experiments 1a and 1b) or prior to (Experiment 2) the experimental session. Despite the success with the processing fluency manipulations, as indicated by faster RTs and higher accuracy in a LDT on the easy items compared to the difficult items, I did not observe an effect of discrepancy either on RM (Experiments 1a and 2) or PM (Experiment 1b).

Instead, I found higher hit rates for easy items relative to difficult items in Experiment 1a, providing support for the fluency attribution account. I also found higher hit rates for low frequency words relative to high frequency words, accompanied by more “remember” and fewer “know” responses (Experiment 2), replicating previous findings in the literature. Even though findings from Experiments 1a and 2 converged with various findings in the literature (e.g., higher hit rates for easy items, Jacoby & Whitehouse, 1989; for low frequency words with more remember responses, Joordens & Hockley, 2000), findings from the three current experiments did not provide any support for the discrepancy attribution account of either RM or PM.

One possibility for not finding the effect of discrepancy in Experiments 1a and 1b was because the expectation of processing fluency built during the experimental session might have been not strong enough for its violation to elicit discrepancy. However, Experiment 2 used high and low frequency words as stimuli. Based on previous findings that showed people judge high frequency words as easier to process (e.g., Begg et al., 1989), and objective measures confirming that judgment (e.g., faster RTs and higher accuracy of LDT for high frequency words compared to low frequency words, e.g., Stanners, Jastrzembski, & Westbrook, 1975), I reasoned that participants would have had pre-experimental expectations for processing fluency of high and low frequency items, just as Whittlesea and colleague’s participants presumably did for words and nonwords or for other stimuli (1998, 2000, 2001a, 2001b). Experiment 2 still found no effect of discrepancy.

Perhaps, the processing fluency manipulation used in Experiments 1a and 1b did not elicit discrepancy because that paradigm did not closely match the paradigms Whittlesea and Williams used (e.g., 1998, 2001b). Considering that the processing fluency manipulation used in Experiment 2 more closely matched the paradigms Whittlesea and colleagues used, perhaps discrepancy was elicited in Experiment 2 but did not influence RM. Unfortunately, in all three experiments of the current project and existing experiments that tested the discrepancy attribution account in the literature, discrepancy is not directly assessed but indirectly inferred by its influence on other measures, such as RM or PM performance. Thus, the current formulation of the discrepancy attribution account (Whittlesea & Williams, 1998; 2000; 2001a, 2001b) cannot distinguish whether the lack of discrepancy effects on RM and/or PM is due to the failure to elicit discrepancy or the failure for discrepancy to be attributed to something, such as familiarity or significance of discrepant items.

Another possibility for not finding the discrepancy effect is that the power could have been insufficient to detect effects in the current project. Taking the relevant experiments reported in the most cited papers of Whittlesea's work, the discrepancy effect seems to be a medium to large size, ranging between Cohen's d of .52 to 2 (Whittlesea & Williams, 1998, 2000, 2001a, 2001b). Cleary et al. (2007) who replicated the discrepancy effect using materials and paradigms from Whittlesea and Williams (1998) reported the effect size of discrepancy in their findings to range between .55 and .65. Based on the findings from Lee and McDaniel (2013) and Thomas and McBride (2015), the effect size of discrepancy in PM seems to range between .53 and .58.

A power analysis was conducted using the smallest Cohen's *d* of .5 from the abovementioned values as an estimated effect size, using G*Power software (Erdfelder, Faul, & Buchner, 1996). Power to detect a medium sized effect (Cohen's *d* of .5; Cohen, 1988) was estimated to be greater than .99 for both Experiments 1a and 2 and .60 for Experiment 1b. Power to detect a large sized effect (Cohen's *d* of .8) was estimated to be greater than .99 for all three experiments in the current project. The result of the power analysis suggests that if my processing fluency paradigm induced discrepancy, there would have been sufficient power to detect a discrepancy effect in Experiments 1a and 2. In addition to the power analysis, the Bayesian analyses reported throughout all three experiments of the current project (except for the hit rates of critical items in Experiment 1a) provided positive support for the null hypothesis, P_{BIC}s $(H_0|D) \geq .90$. Thus, Bayesian analyses also suggest that discrepant items did not differ from nondiscrepant items in RM and PM in the current project.

Given the success with manipulating processing fluency with the paradigms in the current project and the results of the power analysis and Bayesian analysis of Experiments 1a and 2, the lack of a discrepancy effect in my paradigm lends more evidence for the claims that discrepancy and its influence on memory might be more limited than Whittlesea and colleagues have suggested (Cleary et al., 2007; Karpicke et al., 2008). Instead, the significant effect of fluency on hit rates of easy items in Experiment 1a provided additional supporting evidence to the fluency attribution account (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989). The fluency attribution account suggests that fluency can be attributed as familiarity. Faster RTs and higher accuracy of LDT of easy items, compared to difficult items, reflect such fluency. Therefore, higher hit rates of easy items compared to that of difficult items in Experiment 1a are consistent with the prediction posed by the fluency attribution account. It is worth noting, however, that the higher hit rates of low frequency words compared to high frequency words in Experiment 2, when low frequency words had slower RTs and lower accuracy of LDT than high frequency words, are inconsistent with the fluency attribution account as well as the discrepancy attribution account.

Considering that discrepancy is elicited by the mismatch between the expected processing quality and the actual processing quality, the discrepancy attribution account also can theoretically account for higher familiarity for more fluently processed items as long as that fluency somehow creates a mismatch with processing expectations. The critical distinction between the two accounts is that the discrepancy attribution account makes a unique prediction that even less fluently processed items can receive higher familiarity than more fluently processed items as long as less fluently processed items stimulate discrepancy elicited by the mismatch between the expected processing quality and the actual processing quality of those items.

Although Whittlesea and colleagues have provided evidence of less fluently processed items receiving higher familiarity judgments than more fluently processed items (e.g. Whittlesea & Williams, 1998), it is noteworthy to point out that not all of the data supporting the discrepancy attribution account show higher familiarity judgments for less fluently processed items. Some findings showed discrepancy on relatively more fluent items, for which the fluency attribution account could also provide a plausible interpretation. For example, Hansen et al. (2008) interpreted higher ratings of subjective truth of fluent sentences preceded by less fluent sentences with discrepancy being attributed as subjective truth of those sentences. However, the same pattern of result can be explained by fluency attributed as subjective truth of “discrepant” sentences. Considering the difficulty to observe a discrepancy effect across memory paradigms and the potential for an overlap of predictions postulated by the fluency attribution account and the discrepancy attribution account, more caution is needed with using the discrepancy attribution account over the fluency attribution account in interpreting experimental data.

6.1 Implications of Experiment 1b for Discrepancy Effects in PM

Unfortunately, the power analysis found that power was insufficient to detect a medium-sized effect in Experiment 1b. However, the Bayesian analysis provided positive support for the null hypothesis that PM performance for the discrepant PM cues did not differ from that for the nondiscrepant PM cues. Thus, the opposing results of the power analysis and the Bayesian analysis limit the interpretation of the lack of discrepancy on PM performance. A more pertinent factor that may limit the interpretation is the use of nonfocal PM cues. According to the multiprocess framework (McDaniel & Einstein, 2000, 2007), nonfocal PM is supported by strategic monitoring processes that demands limited attentional resources. Experiment 1b used nonfocal PM cues with the premises that (1) the limited attentional resources could make sustaining monitoring throughout the PM task challenging and (2) discrepancy could assist the noticing of nonfocal PM cues that were missed by monitoring. The relatively low PM performance in the nondiscrepant conditions in Experiment 1b confirmed the first premise. The lack of the interaction of PM cue difficulty by nontarget difficulty and the result of Bayesian analysis favoring no effect of discrepancy on PM performance unfortunately cannot confirm or disconfirm the second premise. Perhaps, discrepancy was not elicited and thus could not influence nonfocal PM performance or discrepancy was elicited but did not influence nonfocal PM performance. The current project cannot distinguish between these two possibilities.

Given that the finding from Experiment 1b is inconclusive, it is premature to argue that the findings indicating that discrepancy can enhance focal PM (Lee & McDaniel, 2013; Thomas & McBride, 2015) should be discounted because of the lack of discrepancy effect on nonfocal PM in Experiment 1b. Nevertheless, it is worth considering whether the failure to find effects of discrepancy in RM (in Experiments 1a and 2) undercut the discrepancy-based interpretations for previous reports of enhanced focal PM performance. The generality of the discrepancy effect in RM could be limited as suggested by findings from Experiments 1a and 2. Nonetheless, as suggested previously, the utility of discrepancy on PM performance might be greater than on RM performance because of the lack of explicit prompt for retrieval in PM tasks. If so, the lack of discrepancy effects in RM do not determine if an effect of discrepancy should be observed in PM, regardless of whether the cue is focal or nonfocal. However, if discrepancy does not assist the noticing of nonfocal PM cues and focal PM cues, an alternative explanation is needed for previous reports of enhanced focal PM performance, such as Lee and McDaniel and Thomas and McBride.

One possible alternative is that relatively longer processing time of discrepant PM cues, due to their relative disfluency compared to the nondiscrepant PM cues, in both Lee and McDaniel (2013) and Thomas and McBride (2015) might have enhanced focal PM performance via allowing more time for noticing of PM cues. Indeed, Lee and McDaniel found significantly higher PM performance for PM cues that were the difficult anagrams embedded among easy anagrams than PM cues that were difficult anagrams embedded among difficult anagrams. Based on the RT data, difficult anagrams took longer to process than easy anagrams. PM performance for PM cues that were the easy anagrams embedded among difficult anagrams was only nominally higher than that for PM cues that were the easy anagrams embedded among easy anagrams. Discrepant PM cues in Thomas and McBride were also processed more slowly than nondiscrepant PM cues. Although all participants performed the same ongoing task of a category decision task and received the same PM cues of “arm” and “leg”, the category type for the majority of exemplars presented for the category decision differed between the discrepant and the nondiscrepant conditions. The majority of exemplars were from the “fruit” category in the discrepant condition whereas the majority of exemplars were from the “body part” category in the nondiscrepant condition. Based on Thomas and McBride’s finding that showed the slower RTs of a category decision task for exemplars from the non-majority categories than from the majority category, discrepant PM cues that were exemplars from the non-majority category would have been processed slower, thus, allowing more time for noticing of PM cues, than nondiscrepant PM cues that were exemplars from the majority category.

However, such an alternative explanation that the longer processing time for discrepant PM cues was related to improved PM (rather than discrepancy attribution per se) does not apply to all findings that claimed to show discrepancy-enhanced PM performance. For example, McDaniel et al. (Experiment 1, 2004) either pre-exposed or did not pre-expose nontargets to elicit discrepancy while focal PM cues were pre-exposed in all conditions. They claimed that discrepancy for the pre-exposed PM cues would be greater when nontargets are not pre-exposed than when nontargets are pre-exposed. In support of their claim, they found higher PM performance for the PM cues among nontargets that were not pre-exposed relative to the PM cues among nontargets that were pre-exposed. Higher PM performance in this finding cannot be explained by the longer processing time of discrepant PM cues. Therefore, it is premature to discount the existing findings of discrepancy-enhanced focal PM performance based on the lack of discrepancy effect in Experiment 1b.

Although power to detect a medium-sized effect might have been insufficient in Experiment 1b, future research on nonfocal PM with greater power alone is insufficient to address the issues discussed here. Rather, future research that compares the effect of discrepancy for both focal and nonfocal PM cues can provide more insight on whether or not discrepancy can influence both focal and nonfocal PM. Furthermore, considering the theoretical utility of significance (attributed from discrepancy) in the successful noticing of the PM cue, perhaps discrepancy-driven significance might play a more critical role in a condition that makes the noticing of the nonfocal PM cue challenging. Presenting the first PM cue long after participants received the PM instruction and performed trials of an ongoing task could be such a condition. Scullin et al., (Experiment 4, 2010) reported that when the first PM cue was presented after 500 trials of nontargets, participants were less likely to sustain monitoring processes and only 7 out of 40 participants performed the nonfocal PM task. Such conditions that discourage the engagement in monitoring processes could increase the utility of significance attributed from discrepancy on PM performance.

6.2 Limitations of the Current Project

In addition to the lack of the discrepancy effect on RM and PM, there were additional findings from the current project that are inconsistent with the literature. False alarm rates did not differ between the easy and the difficult items in Experiment 1a or between low and high frequency words in Experiment 2. Also, some patterns of RT results in Experiment 1a and Experiment 2 were puzzling. The RTs of correct recognition judgments on both critical and noncritical items were slower in the conditions with the easy list than in the conditions with the difficult list. There was a main effect of prime type on the RTs of hit responses with no main effect of prime type on the mean accuracy of hit responses in Experiment 2. Below, I discuss the implications of these findings.

Experiment 1a showed comparable false alarm rates between easy and difficult items. According to the fluency attribution account (Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989), the false alarm rate of easy items should be higher than that of difficult items. Experiment 2 did not show the false alarm advantage of low frequency words. Within the current project, it is possible that easy items and high frequency words did not have as many false alarms as easy items and high frequency words did in experiments reported in the existing literature. It is also possible that difficult items and low frequency words had more false alarms than usual. Unfortunately, the current project cannot distinguish which of the two possibilities led to no difference in false alarm rates between the two types of stimuli. .

However, Coane et al. (2011) explained that the high attentional demand of their paradigm of speeded recognition could have decreased the false alarm advantage of low frequency words in their Experiment 1. They reasoned that speeded recognition could have led their participants to rely more on *relative* familiarity (that tracks the increase of familiarity from *absolute/baseline* familiarity) and not on *absolute* familiarity and that high and low frequency did not differ in relative familiarity without a previous study episode. This explanation could account for the lack of a word frequency effect on false alarm rates in Experiment 2 of the current project. Perhaps, a memory test asking for a recognition judgment with a remember/know response as in Experiment 2 or a memory test with a perceptual mask as in Experiment 1a might have been challenging for participants compared to a typical memory test asking for a recognition judgment alone with no perceptual mask.

Experiment 1a showed that the RTs of recognition judgments on both critical and noncritical items in the conditions with the difficult list were faster than the RTs in the conditions with the easy list. There are several possible explanations why such patterns of RTs were observed. Unfortunately, none of those explanations are conclusive. Experiment 2 found faster RTs of hit responses for the recognition probes preceded by the identity primes than by the lexically disrupting primes. Although the faster RTs of hit responses with the identity primes in Experiment 2 were suggestive of the fluency attribution on familiarity judgments, without the higher mean accuracy to match the faster RTs, the implication of the faster RTs was also unclear. One overarching limitation in interpreting the above mentioned RT data stems from the failure to properly partition an RT to the perceptual processes and the recognition processes although Experiment 1a attempted such partitioning by measuring RTs separately for the LDT and the recognition judgment. Future research might consider using more precise measures of RT associated with each of those processes.

6.3 Conclusion

The current project attempted to investigate the generality of the discrepancy effect in RM and PM. Experiment 1a manipulated the processing fluency (easy versus difficult) using the perceptual mask and masked priming. The actual processing fluency was matched or mismatched to the expected processing fluency developed during the experimental session to elicit discrepancy. Although the faster RTs and higher accuracy of LDT of the easy items compared to the difficult items suggested that the processing fluency manipulation was successful, there was no discrepancy effect on recognition memory of items whose expected processing fluency mismatched the actual processing fluency. Instead, in Experiment 1a, higher hit rates of easy items were observed, compared to the difficult items, providing support for the fluency attribution account. Experiment 1b used the same paradigm as Experiment 1a and incorporated a nonfocal PM task to test the discrepancy effect on PM. Despite the possibility that the utility of discrepancy in PM tasks can be greater than in RM tasks, PM performance did not differ across discrepant and nondiscrepant conditions.

Given that the violation of the expected processing fluency developed during the experiment might be insufficient to elicit discrepancy, Experiment 2 attempted to elicit discrepancy by violating the expected processing fluency developed pre-experimentally with high and low frequency words and masked priming. Although the result of the pilot suggested that the priming manipulation was successful, there was no effect of priming or discrepancy on the mean accuracy of recognition judgments. Although I was able to replicate both the hit advantage of low frequency words relative to high frequency words and more “remember” and fewer “know” responses for low frequency words than high frequency words, I did not observe the false alarm advantage of low frequency words. Overall, the patterns of results from the current project raise questions for the generality of the discrepancy effect in RM and PM.

References

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The 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*(1), 199-226.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10(3),* 340-357.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 576-587.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39(3),* 445-459.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610-632.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, *31*(2), 297-305.

Bernstein, D. M., Whittlesea, B. W., & Loftus, E. F. (2002). Increasing confidence in remote autobiographical memory and general knowledge: Extensions of the revelation effect. *Memory & Cognition, 30(3),* 432-438.

Breneiser, J. E., & Mcdaniel, M. A. (2006). Discrepancy processes in prospective memory retrieval. *Psychonomic Bulletin & Review, 13(5),* 837-841.

Bridger, E. K., Bader, R., & Mecklinger, A. (2014). More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. *Neuropsychologia*, *57*, 179-190.

Cleary, A. M., Morris, A. L., & Langley, M. M. (2007). Recognition memory for novel stimuli: the structural regularity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(2)*, 379-393.

Coane, J. H., Balota, D. A., Dolan, P. O., & Jacoby, L. L. (2011). Not all sources of familiarity are created equal: the case of word frequency and repetition in episodic recognition. *Memory & Cognition*, *39*(5), 791-805.

Cohen, A. L., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influences the cost of remembering to remember. *Memory & Cognition, 36(1)*, 149-156.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: the impact of distraction. *Psychology and Aging*, *6*(4), 533-541.

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency X levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition*, *17*(2), 148-162.

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General, 134(3)*, 327-342.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1-11.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680-698.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16(4)*, 309-313.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*(1), 23-30.

Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13(1)*, 8-20.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 21-31.

Goldinger, S. D., & Hansen, W. A. (2005). Remembering by the seat of your pants. *Psychological Science, 16(7),* 525-529.

Greene, R. L. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50(3)*, 259–267.

Guttentag, R. E., & Carroll, D. (1994). Identifying the basis for the word frequency effect in recognition memory. *Memory*, *2*(3), 255-273.

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high-and low-frequency words. *Memory & Cognition*, *26*(5), 951-958.

Guynn, M. J., & McDaniel, M. A. (2007). Target preexposure eliminates the effect of distraction on event-based prospective memory. *Psychonomic Bulletin & Review,* *14(3)*, 484-488.

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology, 44(3)*, 687-691.

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language, 53(3)*, 430-444.

Jacoby, L. L. (1983). Perceptual enhancement: persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(1)*, 21-38.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30(5)*, 513-541.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110(3),* 306-340.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. *Varieties of memory and consciousness: Essays in honour of Endel Tulving,* 391-422.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118(2),* 126-135.

Johnston, W. A., Dark, V., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition, 5,* 3–11.

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*(6), 1534-1555.

Jou, J., Matus, Y. E., Aldridge, J. W., Rogers, D. M., & Zimmerman, R. L. (2004). How similar is false recognition to veridical recognition objectively and subjectively?. *Memory & cognition*, *32*(5), 824-840.

Karpicke, J. D., McCabe, D. P., & Roediger, H. L. (2008). False memories are not surprising: The subjective experience of an associative memory illusion. *Journal of Memory and Language, 58(4),* 1065-1079.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98(2),* 127-140.

Kinoshita, S. (1995). The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition*, *23*(5), 569-580.

Laxon, V., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighborhood size, graphemic changes, spelling–sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 730–748.

Lee, J. H., & McDaniel, M. A. (2009). A study of the discrepancy effect on prospective memory using the sentence-with-a-pause paradigm. Unpublished raw data.

Lee, J. H., & McDaniel, M. A. (2010). Typical PM targets are not typically better than atypical PM targets: Underlying mechanisms of prospective memory retrieval. Unpublished master's thesis.

Lee, J. H., & McDaniel, M. A. (2013). Discrepancy-plus-search processes in prospective memory retrieval. *Memory & Cognition, 41(3),* 443-451.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological*

Review, 87(3), 252-271.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43(3)*, 679-690.

McCabe, D. P., & Balota, D. A. (2007). Context effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 536-549.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 14–24.

McDaniel, M. A. (1995). Prospective memory: Progress and processes. In D. L. Medin (Ed.), *The psychology of learning and motivation (Vol. 33) (pp. 191-221).* San Diego, CA: Academic Press.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieva1: A multiprocess framework. *Applied Cognitive Psychology, 14(7),* 127-144.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field.* Thousand Oaks, CA:Sage.

McDaniel, M. A., Guynn, M. J., Einstein, G. O., & Breneiser, J. (2004). Cue-focused and reflexive-associative processes in prospective memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(3)*, 605-614.

McDaniel, M. A., Shelton, J. T., Breneiser, J. E., Moynan, S., & Balota, D. A. (2011). Focal and nonfocal prospective memory performance in very mild dementia: A signature decline. *Neuropsychology, 25(3),* 387-396.

Pazzaglia, A. M., Staub, A., & Rotello, C. M. (2014). Encoding time and the mirror effect in recognition memory: Evidence from eyetracking. *Journal of Memory and Language*, *75,* 77-92.

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 409–421.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111-164.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition,* 21(1), 89-102.

Rajaram, S., & Neely, J. H. (1992). Dissociative masked repetition priming and word frequency effects in lexical decision and episodic recognition tasks. *Journal of Memory and Language, 31(2),* 152-182.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(2),* 294-320.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4),* 803-814.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes?

Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3), 736-749.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 65-76.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(3),* 347-361.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14*(3), 259-264.

Thomas, B. J., & McBride, D. M. (2015). The effect of semantic context on prospective memory performance. *Memory, (Hove, England),* 1-9.

Wänke, M., & Hansen, J. (2015). Relative Processing Fluency. *Current Directions in Psychological Science*, *24*(3), 195-199.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14(5),* 779-804.

Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(3*), 526-541.

Westbury, C., Buchanan, L., & Brown, N. R. (2002). Sounds of the neighborhood: False memories and the structure of the phonological lexicon. *Journal of Memory and Language, 46,* 622–651.

Whittlesea, B. W. (1993). Illusions of familiarity. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 19(6),* 1235-1253.

Whittlesea, B. W. (2002). False memory and the discrepancy-attribution hypothesis: the prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131(1),* 96-115.

Whittlesea, B. W., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49(1),* 62-79.

Whittlesea, B. W., Masson, M. E., & Hughes, A. D. (2005). False memory following rapidly presented lists: The element of surprise. *Psychological Research, 69(5-6),* 420-430.

Willems, S., Van der Linden, M., & Bastin, C. (2007). The contribution of processing fluency to preference: A comparison with familiarity-based recognition. *European Journal of Cognitive Psychology, 19(1),* 119-140.

Whittlesea, B. W., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98(2),* 141-165.

Whittlesea, B. W., & Williams, L. D. (2000). The source of feelings of familiarity: the discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(3),* 547-565.

Whittlesea, B. W., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 3-31.

Whittlesea, B. W., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *27(1)*, 14-33.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46(3)*, 441-517.

Appendices

Appendix A-1. Word Stimuli Used for Experiments 1a and 1b.

abbey
abroad
absence
abusive
accident
acoustic
acrobat
adequate
admiral
airborne
algae
altitude
anarchy
ancient
angst
apogee
appendix
arbor
Arising
Arsenal
ashamed
autonomy
aviation
awhile
backup
banned
bargain
bearing
becoming
biblical
boundary
brigade
broke
burned
bypass
canine
captive
concise
conclude
covert
crisp
cultural
dairy
deadly
debug
devoted
district
dubious
eagerly
enemy
entropy
expert
exposure
facing
fallen
favored
float
fortune
fragment
funeral
fuzzy
gauge
gesture
gravel
grocery
hail
hanging
harmony
harness
hazard
heal
heating
helmet
historic
hive
horizon
hornet
humanity
humidity
hymn
hypnotic
ignition
ignorant
immunity
imperial
impulse
inch
inherent
initiate
instinct
intent
involve
irony
ivory
jealous
juvenile
knock
labeled
lame
lean
lift
lounge
lust
magical
manage
marker
massacre
metric
midst
modular
moisture
mundane
muscular
naive
nervous
nominal
nonsense
numerous
nylon
obscene
obsessed
officer
omitted
opponent
optical
orbital
ordinary
origin
orthodox
outline
outlook
overload
paddle
peculiar
perceive
protest
proudly
quantum
quiet
recipe
reform
robin
rusty
salmon
satisfy
secured
segment
sewing
shadow
simulate
soil
stable
strategy
subtle
subtract
suppress
swimming
sympathy
tackle
tandem
tangle
tempest
temporal
terrain
testify
thumb
tiger
tired
titan
tobacco
toggle
tomato
tomb
tongue
tooth
torso
tortoise
tough
tourist
toys
tracked
trail
trained
trash
trauma
tribute
trigger
trivia
troop
tutor
typical
ultimate
uniform
urged
usage
voting
wage
wake
warn
wave
wedge
whine
wiring
witness
wizard
wooden
wording
workshop
wormhole
wounded
wrist
yearly
yeast
youth
zenith
zephyr

Note. Words with the initial letter “h” or “o” in Experiment 1a were replaced with other appropriate words in Experiment 1b.

Appendix A-2. Nonword Stimuli Used for Experiments 1a and 1b.

abtrasit
adebi
adophine
adsort
adsotelu
agenka
alumxi
amtigue
ancker
andiety
anomaty
anrosym
antorbal
arcker
areow
argonant
ashorady
atomasy
attifude
atude
auqirec
ausience
ausora
awarkad
barq
batin
begro
beltided
beskat
biot
boiking
breitly
brexe
bulket
butral
catapicy
cenbomed
comtete
consudaf
corepht
courshey
cutsody
daeher
dangimig
deftent
deseft
dycence
dysanny
enudate
etiligle
ettiroty
etuaqe
extosit
fakal
faloubos
fatitue
fondeck
forqing
fukile
gaetic
geronver
giletaec
glecose
gramped
hacper
haelly
hakrid
harvart
hategiae
hattle
hellig
hobrer
hodie
hondimer
honisky
houl
hugorous
hungle
hybrogen
ibiotic
idellya
idvened
ifle
ikentify
imqict
infegral
intility
intly
intreced
inzestor
iplicaet
istued
itolased
kecitan
kingsom
knae
lastine
launfry
lorexy
loying
lufent
lurting
lyfic
mafenist
mangally
markinal
medifoke
midority
mishly
mopive
mosally
motanery
nidely
noce
nodike
nodity
nugula
nuklous
obdly
obgerber
ocune
ogianic
ogion
ondoing
orein
orthid
otenly
otiender
outpoing
outqik
oxiden
padented
Parec
piodeer
pleaked
policked
qouz
quoka
raed
redimene
retisent
retolt
sakity
seithute
serkeant
sethimed
sezerely
singere
sixee
skechic
snouding
squet
strige
striktly
struddle
strunicy
symbotic
tadded
talamt
tarcing
tasfy
tecih
tedonafy
telted
terexue
tenasi
thilk
thoun
tiam
tiffue
toden
toeflert
toel
toileg
tokf
tonshed
toquite
toqure
torbure
tosta
tragide
traik
tramphu
tranth
Traq
trebing
trigent
trothy
tsunala
tursle
ublesife
unfanky
unikom
urgranet
vatation
vatuable
vidible
wabmot
wanerthy
wanger
wealsky
weasher
werended
whape
wikky
worfer
workora
wreish
wure
yaen
yelfow
yerd
Yetl
yielp
zaim

Note. Nonwords with the initial letter “h” or “o” in Experiment 1a were replaced with other appropriate nonwords in Experiment 1b.

Appendix B-1. Materials Used for Experiment 2: High Frequency Words.

Stimulus	Prime	Stimulus	Prime	Stimulus	Prime
affect	effect	firm	fire	rice	race
appeal	appear	fork	work	road	read
ball	call	fort	sort	rose	lose
batch	watch	header	leader	sake	take
beach	reach	hero	here	sea	set
bear	hear	home	some	sector	vector
beast	least	hood	food	sheep	sleep
belt	felt	hope	home	shirt	short
blade	blame	horn	born	side	site
book	look	jacket	packet	sight	night
bottle	battle	king	kind	singer	finger
bread	break	lake	take	snake	shake
bull	pull	lamp	camp	socket	rocket
cage	page	lane	line	sound	found
cancer	cancel	laser	later	stage	state
car	far	lead	head	suffer	buffer
cat	cut	letter	better	sun	run
cell	tell	life	like	threat	thread
chance	change	loss	less	town	down
clay	play	lung	long	trail	train
close	clone	mall	male	tree	free
cold	hold	math	path	truck	track
cow	low	meat	beat	vice	nice
creek	greek	mood	wood	wait	want
cup	cut	mouth	month	wealth	health
dean	mean	name	same	white	while
desk	disk	node	mode	wine	mine
dress	press	peace	place	winner	winter
fate	rate	plane	place	wish	with
father	rather	pool	cool	world	would
fear	hear	reed	feed	yard	hard
fire	fine	retail	detail	zone	none

Appendix B-2. Materials Used for Experiment 2: Low Frequency Words.

Stimulus	Prime	Stimulus	Prime	Stimulus	Prime
adept	adopt	gravel	travel	realty	really
altar	alter	grief	brief	rind	wind
arid	grid	groom	gloom	roast	coast
bead	bear	harp	harm	romp	ramp
beaker	weaker	haste	taste	rude	rule
bean	mean	hearth	health	rumor	humor
beet	bent	hermit	permit	seam	seat
blaze	blame	hoof	roof	shaker	shakes
bleak	break	hose	rose	silo	silk
bleed	breed	hurl	hurt	sinus	minus
boar	boat	isle	idle	soak	soap
boarder	boarded	jade	made	soar	sour
broom	bloom	keg	leg	spice	spite
bruise	cruise	kilt	tilt	spoon	spool
cavern	tavern	kneel	knees	steed	steel
cheat	cheap	lender	gender	stilt	still
cheer	sheer	loft	lift	stool	spool
cloak	clock	lymph	nymph	stroll	scroll
comb	bomb	mayor	major	tease	cease
coral	moral	meek	seek	toad	load
crease	create	mesh	mess	torch	touch
dense	sense	munch	bunch	tuba	tube
dread	bread	nail	fail	tunic	tonic
dune	tune	niece	piece	valve	value
dusk	duck	olive	alive	vase	ease
earring	earning	otter	outer	vine	nine
flask	flash	owl	oil	vulture	culture
flea	flew	peach	teach	weave	leave
fowl	bowl	plum	plug	wheat	cheat
frail	trail	poke	joke	wreath	breath
fright	bright	polar	solar	yearn	learn
gaze	gate	quill	quilt	zoo	too