1998

Changes in environmental context and the mirror effect in recognition memory

Craig Andrew Livermore
Wilfrid Laurier University

0-612-33805-3

Changes in Environmental Context and the Mirror Effect in Recognition Memory

By

Craig Andrew Livermore

B.Sc. (Hons.), Wilfrid Laurier University

THESIS

Submitted to the Department of Psychology

in partial fulfilment of the requirements

for the Master of Arts Degree

Wilfrid Laurier University

1998

Abstract

The present study examined the effect of environmental context on the mirror effect in recognition memory. In seven experiments, participants studied either high (HF) and low-frequency (LF) or noun and nonnoun pairs of words followed in a old/new item recognition test by which proportion correct, response time, and confidence judgments were measured. Single item or word pair targets and distractors were presented in same- or different-context conditions. Context was defined as the unique combination of foreground and background colour and position on a computer screen for two experiments while position was removed as a context variable for five experiments due to influences on response time. The nature and types of the different-context conditions were manipulated in addition to the number of study context repetitions. Single item recognition tests produced statistically independent mirror effects (higher hit rates and lower false alarm rates for LF compared to HF words and nouns compared to nonnouns) and effects of context (higher hit rates and higher false alarm rates for old compared to new contexts). Results also indicated that words were not associated with specific contexts but rather the type of context as new combinations of old context elements and repeating old contexts had no effect on performance. Confidence level also had no effect on performance as these results were replicated when conditionalized on high-confidence responses. Tests of word pairs failed to show an effect of context while demonstrating an effect of word type for hits only. For the most part, results support the view that effects of environmental context are due to a response bias arising from changes in response criteria (upon which recognition decisions are based) across test context conditions. This leads to a greater likelihood of "new" responses when recognition tests are in a novel context and a greater likelihood of "old" responses when recognition tests are in a same/old context. Evidence of differences in discriminability between different word types supports the

notion that the mirror effect is a memory-based phenomenon characterized by properties that make participants more sensitive to LF over HF words (and nouns over nonnouns).

Acknowledgements

I would like to take this opportunity to recognize those who have proven to be invaluable in my pursuit of this research and my academic career at Wilfrid Laurier University. Most importantly, I wish to thank my undergraduate and graduate thesis advisor Dr. William Hockley. Dr. Hockley has been a ceaseless fount of support and inspiration for me during my tenure in Psychology at W.L.U. May this research project serve as a tribute to his incredible ability to bring out the best in his students. I also wish to express sincere gratitude to my committee members Dr. Rudy Eikelboom and Dr. Roger Buehler for their insightful suggestions and valuable assistance over the course of this project. Finally, I wish to give thanks to the many professors and fellow students that have made my time at Laurier both personally and intellectually rewarding.

Table of Contents

List of Tables

List of Figures

Changes in Environmental Context and the

Mirror Effect in Recognition Memory

Within the realm of studies on human memory, there has been much interest in the effects of context, more specifically environmental context, in tests of recall and recognition. Murnane and Phelps (1993) identify environmental context as any information that is present in the encoding/processing environment, yet is incidental to the cognitive task used during learning. Changes in environmental context from encoding to retrieval have been indicated as an important component of the process of forgetting (Smith, 1988). Until recently, it was believed that memory tests involving recall of information are greatly affected by changes in context from learning to test but memory tests involving recognition of information remain relatively unaffected (Godden & Baddeley, 1975, 1980; Smith, 1988). However, more recent studies (Murnane & Phelps, 1993, 1994, 1995) have demonstrated that recognition memory can also be affected by changes in environmental context. The present study focuses on tests of recognition memory, about which several theories have been proposed to account for these varying recognition study results.

Many current views concerning memory attempt to explain the mechanisms behind how an event or episode is retrieved from memory with reference to the encoding specificity principle (Tulving & Thompson, 1973). This principle holds that the ability to retrieve an episode from memory is related to the degree to which the information at retrieval matches that present during encoding. Context information plays an important role in this process. As stated above, context is seen as any information present during the encoding task that is not integral to the task itself. On subsequent retrieval tasks, providing the same information that was present during encoding tends to increase the ability to retrieve the item or episode of interest. A recent method used to manipulate environmental context (i.e. details of the surrounding environment) in studies of context effects is via computer (Murnane & Phelps, 1993, 1994, 1995). In these studies, to-be-remembered (TBR) items were presented on a computer screen and the combination of foreground colour, background colour, screen location, and type font was used as the environmental context. This way, several different environmental contexts can be used with relative ease. These studies are of particular interest to the current study.

The idea behind tests for environmental context effects is quite simple. It should be expected that changing the environmental context of a TBR item from learning to test will decrease it's chances of being retrieved leading to a reduction in performance on different-context items compared to items that are presented in the same context as at encoding. While studies have determined that changes in context from study to test do in fact reduce performance on tests of recall (e.g. Godden & Baddeley, 1975), studies concerning context-dependent recognition have yielded less consistent results (Smith, 1988). Some research has shown recognition memory to be unaffected by changes in environmental context (e.g. Godden & Baddeley, 1980; Smith, Glenberg & Bjork, 1978). Other studies, however, have shown that testing in an environmental context that differs from that present during learning affects recognition memory (Murnane & Phelps, 1993, 1994, 1995).

Murnane and Phelps (1993, 1994, 1995) have attempted to explain the discrepancies in findings on the effect of environmental context changes on recognition memory. They proposed a global activation theory of memory as a framework for examining this phenomenon. This theory holds that recognition decisions are contingent upon the activation of a (possibly) large set of items in memory and are not based upon the representation of the single test item in memory. The global activation model posits that the degree of activation for a particular item in memory is determined by the match between information present in the retrieval cue and the information stored in memory. This summation of memory activation provides the basis for recognition decisions. To take this idea further, both targets (items that were learned at encoding) and distractors (items not present during encoding) would have varying degrees of activation during a test of recognition depending on the information provided by the environmental context. For instance, a target that is presented at test in the same context as it appeared in during learning would have a high degree of activation from the matching of context and item information. On the other hand, a target that is presented in a context that differs from the context used during learning will lose the activation that is provided by the matching of contexts and will have only the matching of item information as a basis for recognition. Distractors that are shown in a context that was seen during learning would not match on item information, but would have a significant degree of activation provided by the matching of contexts. Distractors that are presented in a different context than was seen at study would then have neither matching item nor context information.

Using the global activation approach, Murnane and Phelps (1993) argued for an important distinction between studies that have found that changes in context do not affect recognition memory (e.g. Godden & Baddeley, 1980; Smith et al., 1978) and their own experiments that demonstrated context effects in recognition memory. The difference lay in the design of the experiments. The above studies that did not reveal context effects manipulated test context between subjects. It was believed that, if recognition memory decisions are based on a global activation view, a between-subjects design could disguise the effects of a different-context test. The global activation theory states that activation for both target items and distractors will be lower in a different context because the context information will not match from the cue to the representation in memory. Assuming that participants use criterion settings to optimize correct recognition for targets ("hits") and reduce incorrect recognition for distractors ("false alarms"), it is believed that these criterion settings will differ between groups in a same-context and different-context condition. However, when test context is varied within subject lists, the response criterion will be less likely to change with context changes. Using a within list design, Murnane and Phelps (1993) found a consistent environmental context effect; that is, same-context targets were recognized better than novel-context targets and novel-context distractors were recognized better than old-context distractors. This study showed the importance of using a within subject and within list design when testing recognition memory.

In a further study, Murnane and Phelps (1994) discussed the potential difficulties involved in a different type of testing procedure. They compared two types of recognition testing procedures, one in which different-context items were presented in a completely novel context and one in which different-context items were presented in a context that had been seen during learning but was not paired with that particular item. It was found that recognition tests involving an item in a context that was used during learning but was different from the context in which that item had originally been tested with either eliminates or drastically reduces environmental context effects that would normally be seen if items are tested in a completely novel environmental context. These findings provide an additional reason why some studies have failed to find context effects in tests of recognition memory as they suggest that a context a participant is familiar with cannot serve as an effective "different" context and stress the importance of using a completely novel context at test.

In yet another study concerning environmental context changes in recognition memory, Murnane and Phelps (1995) looked at the relative strength of the memory cue used in the test for recognition memory. They discuss two possible predictions concerning memory strength and context effects in recognition. Smith (1988) introduced the first as the outshining hypothesis. This hypothesis predicts that as the strength of the TBR items increases, context effects decrease (given that the strength of the context cue is held constant). This hypothesis explains studies that have not found context effects by arguing that their results are caused by differences in relative test context versus test item cue strengths (Murnane & Phelps, 1995). However, as Murnane and Phelps pointed out, current global activation theories predict that as item strength increases, then context effects will either increase or remain the same. Over three experiments, Murnane and Phelps manipulated item strength through spaced repetitions (1 to 5 presentations), study time (1 to 5 seconds), and orienting task (semantic vs. graphemic). The overall pattern of results favored current global activation models as context effects were found for both targets and distractors at all levels of strength, and the size of the effect either increased or remained unchanged as item strength increased. If the size of the effect had only remained unchanged then it would not be possible to discount the outshining hypothesis as the manipulations of item strength may not have been effective; however, it is difficult to explain the increase in the context effect via the outshining hypothesis.

It should be noted, however, that the argument posed by Murnane and Phelps (1993) is not without criticism. With their global activation approach, they argue that, at test, an item's strength or familiarity and the familiarity of an old context combine so that the mean of the underlying strength or familiarity distribution for old items in an old context will be higher than the mean strength of the distribution of old items in a new context. Likewise, the mean of the distribution for new items in an old context will be higher than the mean of the distribution for new items in a new context. Murnane and Phelps argue that this strengthening together with a single decision criterion produces the higher hit rate and false alarm rate for old and new items respectively tested in an old context.

However Murnane and Phelps (1993) also found that d', the index of discriminability, does not generally change due to context changes. It could be argued that participants change their decision criterion in such a way that the decision criterion is lower for test items (old and new) presented in an old context compared to test items presented in a novel context. Such a criterion change would also produce higher hit and false alarm rates for old and new items tested in an old context. Even using a within-subject, within-list design, participants could still use different criteria for old and novel contexts because the old and novel contexts would be apparent to the participants on each test. This possibility is thoroughly considered by Feenan and Snodgrass (1990). Their explanation of the effects of changing context from learning to test suggest that differences in performance are primarily a result of a response bias and less a result of differential sensitivity to new and old items. According to this view, participants adopt a more conservative response criterion when there is change in context, leading to a lower hit rate for old words but a higher correct rejection rate (i.e. a lower false alarm rate) for new words. Feenan and Snodgrass reviewed several other studies to demonstrate that changing context does not affect a participant's ability to discriminate between an old and a new word but rather creates a bias that makes participants less likely to identify a word as old if it is presented in a novel context.

The effects of context on item recognition in terms of changes in the hit rate and false alarm rate between old and novel context conditions have been thoroughly documented (e.g. Feenan & Snodgrass, 1990; Murnane & Phelps, 1993, 1994, 1995). Another phenomenon of recognition memory, the mirror effect, is also seen in changes of hit and false alarm rates. The mirror effect is evidenced when items of different classes (e.g. high- and low-frequency words) are learned and subsequently tested, stimulus classes that are more correctly identified as old (previously seen in a study list) when they are old are also more correctly identified as new when they are new. Item types that are recognized poorly as old when they are old are also poorly recognized as new when they are new. In short, performance on new and old items mirror each other (Glanzer & Adams, 1985).

A variety of different stimulus manipulations have been shown to produce the mirror effect (see Glanzer & Adams, 1985 for a review). Hockley (1994) used comparisons of nouns vs. nonnouns, high-frequency vs. low-frequency words, and high vs. low word concreteness to examine this effect. He found that a mirror effect was reliably produced in item recognition for all three stimulus manipulations. In addition, Hockley found that patterns of response latency also showed the mirror effect for two of the stimulus manipulations. So not only did participants respond more accurately to items of the more memorable class, they also responded faster.

The mirror effect has been difficult to explain in terms of the majority of current models of recognition memory (Hintzman, Caulton, & Curran, 1994). Specifically, most of these models falter in that they attempt to describe recognition memory in terms of signal detection theory. In this theory, new and old items are believed to vary along a single dimension of memory strength (e.g. Gillund & Shiffrin, 1984). With signal detection theory, it is difficult to explain how a more recognizable stimulus class can have a greater mean strength than a less recognizable stimulus class when items are old, but a lesser mean strength when items are new.

However, several theories have been proposed to account for the mirror effect in recognition memory, the most evolved being those dealing with the notion of rescaling (e.g., Glanzer & Adams, 1990; Glanzer, Adams, & Iverson, 1991; Kim & Glanzer, 1994), which basically proposes that old and new item distributions for different stimulus classes are properly ordered before a recognition decision is made. Glanzer and his colleagues have argued for an attention-likelihood theory, a detailed rescaling approach. According to the attention-likelihood theory, in a typical yes/no recognition task, participants generate a likelihood ratio to define a decision criterion based on the original strength judgment of the test item. The likelihoods of the test item coming from or not coming from the study list are compared based on the number of marked features a studied or unstudied item is likely to share with the test item. This ratio is what is used to determine whether the individual strength judgment (of the test item) comes from the new item distribution or the old one. The mirror effect occurs as a result of likelihood ratios being calculated for each stimulus presented since these ratios will differ for items from different stimulus classes.

Hirshman and Arndt (1997) compared two accounts of the mirror effect; a memory-based view and a decision-based view. According to the memory-based account, the mirror effect results from certain types of items being characterized by properties that distinguish them from items in memory. For example, while low-frequency old words have a greater memory strength than high-frequency old words (due to participants' greater ability to remember these items), the opposite occurs for new items due to the fact that low-frequency new words have properties that better distinguish them from items stored in memory (from the study list) than high-frequency new items. Hirshman and Arndt suggest that low-frequency items could have orthography or phonology that discriminate them from items in memory more effectively than high-frequency words. The separation of these old and new distributions on either side of a single response criterion is then what produces the mirror effect.

Hirshman and Arndt (1997) also discussed a decision-based explanation that suggests that all new items are represented in a single distribution (that is lower in memory strength than the old distributions) and the mirror effect that is based on changes in response criterion for low- and high-frequency words. More specifically, the greater memory strength of low-frequency items creates a higher decision criterion than does that of high-frequency items. This upward shift in the decision criterion is such as to not compromise the higher hit rate for LF items but produce a lower false alarm rate (i.e. higher correct rejections) for low-frequency test items. Hirshman and Arndt tested these two theories based on the assumptions each made about the independence of hit and false alarm rates. The memory-based view holds that differences in hits and false alarms occur independently (i.e. the separation between the new distributions that produces false alarm differences is based upon a different factor than the separation between the old distributions producing hit rate differences). However, the decision-based view suggests that differences in hit and false alarm rates are dependent. That is, the separation of the old distributions responsible for hit rate differences causes differential shifts in response criteria that produce false alarm differences. Over several experiments, Hirshman and Arndt demonstrated that hit and false alarn rates associated with word frequency and word concreteness can be independent of one another, supporting the memory-based view of the mirror effect over the decision-based view.

The present study was designed to investigate both environmental context and the mirror effect together. As has been discussed, both effects include changes in the hit rates and false alarm rates. However, although each effect leads to an increase in hit rates (i.e. same context conditions and low-frequency items both do this), false alarm rates go in opposite directions for each effect (i.e., false alarm rate decreases for low-frequency items but increases in a same context condition). Therefore, these two effects are two quite different phenomena.

A major question of the present experiment is whether or not the context effect and the mirror effect are independent effects as the answer has important implications for the nature of these effects. For example, Feenan and Snodgrass (1990) have suggested that the effects of changes in environmental context are a result of response bias. Conversely, Hirshman and Arndt (1997) have provided support for a memory-based explanation of the mirror effect. Statistical independence of these two effects would lend support to these two separate views. However, the presence of an interaction between these two manipulations would suggest that the mirror effect and/or the effects of context are a result of both response bias and differences in discriminability.

One obvious prediction is that they are indeed independent effects. Hintzman et al. (1994) have suggested that the mirror effect may be best described as a fundamental or intrinsic aspect of recognition memory. If this is indeed the case, it would be expected that the mirror effect would not easily be affected by other types of variables. It is also possible to combine the global activation model of Murnane and Phelps (1993, 1994, 1995) with the attention-likelihood model of Glanzer and his colleagues (Glanzer & Adams, 1990; Glanzer et al., 1991; Kim & Glanzer, 1994) to predict that context would in fact add to the strength of the underlying stimulus distributions which are then rescaled to produce the mirror effect.

It is also possible that these two effects could interact. Support for this prediction can be found from a study by Smith et al. (1978) who were interested in determining whether context affects the semantic encoding of words. They predicted that high-frequency words would be more affected by changes in environmental context than low-frequency words on a later recognition test. It was believed that high-frequency words were more likely to be encoded differently in different contexts and would therefore be influenced more by changes in environmental context. The general trend that was found was that low-frequency words were not affected by changes in environmental context while high-frequency words were recognized more poorly in a different environment. This interaction was found to be small but significant in one experiment but not in another that attempted to control the semantic sense in which the word was encoded. According to the work of Murnane and Phelps (1993), two aspects of Smith et al.'s procedure may have affected their results. First, Smith et al. used a between-subjects design. Murnane and Phelps discussed the possibility of differential response criteria playing a role in between-subjects designs so the lack of a strong context effect may be a result of this. Second, Smith et al. used only one context during their learning session so words were either presented in the same single context as before or in a new one. The difficulty with this is the possibility that participants may mentally reinstate the learning environment as a retrieval strategy thereby reducing or eliminating context effects. Murnane and Phelps (1993, Experiment 2) found that when mental reinstatement was suggested to participants, context effects were found for multiple learning context conditions but not for a single learning context condition. Additionally, Smith et al.'s changed context condition may have appeared familiar to participants and would therefore have been insufficiently novel to elicit the effects of changing context (much like the AB-A condition used by Murnane & Phelps, 1994; to be discussed in the introduction to Experiment 1).

The following experiments were designed to explore the joint effects of manipulating context and word type. Experiments 1 through 5 used the stimulus manipulation of low-frequency versus high-frequency words to demonstrate the mirror effect. Experiment 1 introduces the procedure of presenting low-frequency (LF) and high-frequency (HF) word pairs in varying contexts (foreground colour, background colour, and screen position) and then employed a recognition test of single LF and HF words in the same-old context, novel context, or a rearranged context (composed of old context elements in a novel combination). Experiment 2 used intermediate contexts at test (i.e. same-old screen colours but new location and new screen colours but same-old screen location) to assess how varying the level of context similarity from study to test affects recognition performance. Experiments 3A and 3B simplified the design to only same-old and novel text context conditions and removed screen location as a context parameter in an attempt to clarify response time data. The role of confidence level was investigated in Experiment 4 by having participants rate their level of confidence in their recognition decisions. Experiment 5 used single and multiple presentations of contexts during study to determine if increasing the strength of a study context increases the magnitude of the effect of changing context. Experiments 6 and 7 used the stimulus manipulation of nouns versus nonnouns to demonstrate the mirror effect. Experiment 6 replicated Experiment 3B using nouns and nonnouns in order to determine if the previously observed results would generalize to another stimulus manipulation used to produce the mirror effect. The effects of changing context on the mirror effect for associative memory was investigated in Experiment 7 by presenting word pairs at test instead of single words.

Experiment 1

The purpose of Experiment 1 was to extend previous research on context-dependent recognition. More specifically, the effects of changing context on the mirror effect was investigated. Using word frequency as the stimulus variable, it could be expected that changing context from learning to test would hinder recognition performance for high-frequency words (normally recognized less well as new or old) while leaving low-frequency words relatively unaffected. In other words, performance would be expected to be diminished for high-frequency items when context is changed from learning to test while performance on low-frequency words would remain unchanged. Therefore, while it would be expected that the mirror effect (both accuracy and response latency measures) would be found for items tested in the same context as experienced during learning, the mirror effect should be more pronounced for items tested in a different-context (the difference between performance on low- and high-frequency words would be greater) if the effect of word frequency is congruent with, or can be represented as, a difference in strength. These predictions are consistent with those made by Smith et al. (1978) and are in agreement with the outshining hypothesis that predicts that as item strength increases (represented by low-frequency words), context effects will decrease (Smith, 1988).

However, global activation models would predict that as item strength increases, the magnitude of the context effect will either increase or remain the same. This tends to suggest that changing context would result in a greater detriment for low-frequency as opposed to high-frequency words. Such a finding would provide support for global activation models of memory. Additionally, global activation models could also predict no effect of changing context given the mirror effect. The predictions vary with the specific model and whether these models use activation functions that produce divergent or parallel high and low context match functions (Murnane & Phelps, 1995).

In an attempt to help distinguish whether effects due to context or word frequency are due to differences in discriminability or changes in response criteria, d', A' and B''_D scores were used as dependent measures [1]. If effects of context are due to differing response criteria (as is suggested in Feenan & Snodgrass, 1990), differences in B''_D scores would be expected between context conditions. Additionally, if Hirshman and Arndt (1997) are correct in their proposal of the mirror effect being a result of a memory-based process over a decision-based process, differences in d' and A' would be expected across low- and high-frequency words.

A second purpose of this experiment deals with the nature of the "different" context. Murnane and Phelps (1994) found that when items are tested in a different context from that in which they were learned, but the context itself had been seen during the learning phase, context effects did not occur. In their experiment, effects of changing context were found when a recognition test took place in a novel context that was not seen previously, but were not found when testing took place in a context that was from the study phase but was not originally paired with the item. In other words, as long as a context had been seen during study, it led to better performance than novel context test items whether the test items had actually been presented in that context or not. For this reason, they stressed the use of a completely novel "different" context test condition (i.e. consisted of a combination of screen colours and location that not only differed from the study context but was never seen during study) in order to obtain reliable context effects in recognition memory. Experiment 1 used two "different" test contexts; a context that had a completely novel appearance, and a context that used a rearrangement of previously seen context elements to create a new (not seen before) context. This differs from Murnane and Phelps' AB-A procedure in that Experiment 1's rearranged contexts are contexts not used during study while the AB-A procedure simply presented items in different context from the same study list. If a context effect occurs in both "different" contexts, then it could be said that a unique pattern of context elements is sufficient to produce context effects. However, if a context effect is produced only in the novel test context, then there would be support for the idea that no part of the surrounding context can be effective in producing a context effect if it has been previously seen, possibly because the distinction between a different context and a same context condition is harder to make as the similarity of the contexts increases.

Similar to Murnane and Phelps (1993, 1994, 1995) the present study defined context as a unique combination of foreground colour, background colour, and screen location. In Experiment 1, participants viewed pairs of high-frequency and low-frequency words presented in various contexts. Following the procedure used by Murnane and Phelps, participants were instructed to associate the two words of each pair in order attenuate the chance that words learned in one context would become associated with another learning context. This procedure allows for the assumption that information about individual words and their associations are adequately and similarly stored in memory. During the test phase, participants were given an old/new recognition test for these words and distractors (both high-frequency and low-frequency) in either the same or different contexts. The different context was either completely novel or a rearrangement of previously seen context elements. That is, the "novel" context was a completely novel context, composed of completely unseen elements (colours and location not seen during learning) while the "rearranged-old" context was simply another unique combination of context elements (each of which had been seen in a previous context combination but with different context elements).

Method

Participants. Eighty Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit for participation.

Stimuli and Apparatus. Study pairs and distractors were randomly selected without replacement for each subject session from Glanzer and Adam's (1990) pool of 248 high-frequency words (mean frequency of 177.3 per million, according to Kucera & Francis, 1967) and 248 low-frequency words (mean Kucera-Francis frequency of 11.2). List generation, display, and response recording was controlled by IBM-compatible laboratory computers equipped with colour monitors. Participants used the "z" and "/" keys to respond. The keyboards were fitted with black plastic covers that only exposed the two response keys.

Context elements consisted of 4 different possibilities for foreground colour, background colour, and screen positions. The background colours consisted of yellow, pink, purple, and orange. Foreground colours consisted of blue, green, red, and magenta. Screen locations were the top right-hand, top left-hand, bottom right-hand, and bottom left-hand corners of the monitor screen. Thus, there were 64 unique combinations of the different context elements. The elements for the completely novel "different" context condition were white text on a black background with a centre screen position.

Procedure. The study consisted of one experimental session. Each session was divided into 6 learning/test trials. At the beginning of the experimental session, each participant was seated in front of a computer terminal and provided with written instructions for the experiment (see Appendix A for instructions to participants). Participants initiated a trial by pressing either response key. During the study phase, participants viewed word pairs presented on the computer screen in different learning contexts (each word pair was presented in its own context). Participants were instructed to try and relate the words in each pair as an effective technique for remembering information. There were 24 word pairs presented in total for each trial. Word-frequency was manipulated as a within-subjects factor. Items that made up the two members of each word pair were either high-frequency (HF) or low-frequency (LF) nouns. Each study list consisted of half HF and half LF word pairs. Each pair was presented for 3 seconds followed by 500 ms of blank screen between word pair presentations. Each word pair at study was presented in its own context, consisting of a unique combination of a foreground colour, background colour, and screen position. The first and last four word pairs were not tested and served as buffers for the effects of primacy and recency.

After presentation of the last word pair of the learning phase, participants received a 5 second signal indicating the commencement of the test phase. The test phase consisted of a 32 single-item, "old-new" recognition test for the words from the learning phase. One word from each of the 16 study word pairs (excluding the first and last four word pairs) was tested. These 16 targets were drawn randomly from study pairs with the constraint that they were equally divided between the first and second members of the study pairs. One half of the targets were high-frequency words and one half were low-frequency words. In addition, there were 16 distractors (new words) in the test phase. One half of the distractors were high-frequency words and one half were low-frequency words.

Test context was manipulated within lists. Half of the targets from each list were tested in their learning context (same-old context condition) and the remaining half of the targets were tested in a different context that was not seen during the learning phase. Different-context conditions were of two forms. The first was a completely novel context (novel context condition) consisting of screen colours and positions that were not presented in the study phase. The second different-context condition (rearranged-old context condition) was another unique combination that was not present in the learning condition combinations while each of its components (screen colours and locations) had been involved in other learning contexts. Half of the different-context presentations were in each form. An equal number of HF and LF targets were tested in same- and different-test contexts. Half of the distractors were tested in one of the learning contexts (same-context distractors). The remaining distractors were tested in the novel or rearranged contexts (different-context distractors). Presentation order of lists, study items, and test items were randomized for each participant.

Upon presentation of each test item, participants were asked to respond whether the test word was "old" (present in the learning phase) or "new" (not present in the learning phase). Participants responded by pressing one of two keys on the computer keyboard designated for "old" and "new" responses. The latency of the responses was measured from the onset of the test probe to the key press. The test probe was removed from the screen after the response and the next test probe was presented after a 1 sec interval.

After completion of the test phase, participants were given a message indicating that they could start the next trial when they were ready by pressing any key. This procedure continued for 6 trials.

Design. A 2 x 3 within-subjects factorial design was used for this experiment. The two factors were word frequency (high vs. low) and type of context (same-old vs. rearranged-old vs. novel).

Results and Discussion

Analyses were conducted with mean hit rate (correctly identified targets), mean correct rejection rate (correctly identified distractors), and d', A' and B''_D scores as dependent measures. Additionally, mean response time (RT) for hits and correct rejections were also analyzed. Means were calculated for each subject for each condition.

Figure 1 displays the proportion of correct responses for new (i.e. correct rejections) and old (i.e. hits) items for LF and HF words in each context condition. The proportion of correct responses are arranged in Figure 1 so that the mirror effect is evidenced by a U-shaped pattern of results for LF and HF words on both "new" and "old" items. Effects of context condition are shown by higher performance in the novel context condition on "new" tests and lower performance in the novel context condition on "old" tests. A 2 X 3 analysis of variance

(ANOVA) based on hit rate showed a highly significant effect of word frequency [F(1,79) = 72.4, MSe = .013, p<.001] and a significant effect of context condition [F(2,158) = 21.9, MSe = .020, p<.001]. However, there was no significant frequency by context condition interaction [F(2,158) < 1, MSe = .010]. Paired sample t-tests for hits collapsed across word frequency showed a significant difference between the proportion correct in the novel context condition and the same-old context condition [t(159) = 6.243] and between the novel context condition and the rearranged-old context condition [t(159) = 6.216], but no significant difference between the same-old and rearranged-old context condition [t(159) = .183].

Insert Figure 1 about here

The ANOVA for correct rejection rate also showed a significant effect of word frequency [F(1,79) = 24.1, MSe = .010, p<.001] and a significant effect of context condition [F(2,158) = 20.2, MSe = .013, p<.001]. Similar to hits, there was no significant frequency by context condition interaction [F(1,158)<1, MSe = .009] for correct rejections. Paired sample t-tests for correct rejections collapsed across word frequency revealed a significant difference between the proportion correct in the novel context condition and the same-old context condition [t(159) = 5.346] and between the novel context condition and the rearranged-old context condition [t(159) = 5.834], but no significant difference between the same-old and rearranged-old context condition [t(159) = 1.299].

Mean d', A', and B''_D values are presented in Table 1 for HF and LF words in the new, same old, and rearranged old context conditions. A 2 X 3 ANOVA was conducted for each of the three values as dependent measures. Analysis of d' showed a significant effect for frequency

[$F(1,79) = 72.5$, MSe = .53, $p<.001$]. The main effect of context condition [$F(2,158) = 2.96$, MSe = .46, ns] and the frequency by context interaction [$F(2,158)<1$, MSe = .51] did not approach significance. Similarly, analysis of A' showed a significant effect of frequency [$F(1,79) = 50.33$, MSe = .008, $p<.001$] while the main effect of context condition [$F(2,158)<1$, MSe = .005] and the frequency by context condition interaction [$F(2,158) = 1.33$, MSe = .005, ns] did not approach significance. The analysis of B''_D (the estimate of the decision criterion associated with A') showed a significant effect of frequency [$F(1,79) = 5.34$,MSe = .17, $p<.01$] and a significant effect of context condition [$F(2,158) = 37.3$, MSe = .22, $p<.001$]. The frequency by context condition interaction for B''_D did not reach significance [$F(2,158) = 1.53$, MSe = .16, ns]. Both d' and A' were calculated for this experiment though A' has been shown to be a slightly better measure of discrimination in situations where criterion changes occur (Donaldson, 1993). However, even though evidence was found to suggest criterion changes in the present study, the pattern of A' and d' was similar.

Insert Table 1 about here

The mean RTs for correct new and old responses in each condition are displayed in Figure 2. A 2 X 3 ANOVA was performed separately for hits and correct rejections with mean RTs for hits as the dependent measure and with mean RTs for correct rejections as the dependent measure. The analysis of the mean RTs for hits revealed a significant main effect of frequency [$F(1,79) = 16.9$, MSe = 119,126.3, $p<.001$] but no main effect of context condition [$F(2,158)<1$, MSe = 143,521.9] and no frequency by context condition interaction [$F(2,158)<1$, MSe =

105,059.7]. These results are seen in Figure 2 as a faster RT for LF targets than HF targets and no difference between the context conditions within the different frequencies.

Insert Figure 2 about here

The analysis of the mean RTs for correct rejections revealed a significant effect of context condition [$F(2,158) = 37.3$, $MSe = 75{,}797.1$, $p<.001$] but no effect of frequency [$F(1,79) = 2.17$, $MSe = 100{,}707.9$, ns] and no frequency by context condition interaction [$F(2,158) = 1.82$, $MSe = 69.445.2$, ns]. Paired sample t-tests for correct rejections collapsed across word frequency showed a reliable difference between the RTs in the novel context condition and the same-old context condition [$t(159) = 8.399$] and between the novel context condition and the rearranged-old context condition [$t(159) = 7.189$]. However, there was no significant difference between the same-old context condition and the rearranged-old context condition [$t(159) = .363$]. These results are evidenced by the faster RT for "new" items presented in a novel context compared to same- and rearranged-context items for both LF and HF words, but no difference between LF and HF items.

Experiment 1 investigated the mirror effect in recognition memory and sought to examine the effects of changing environmental context from study to test on HF and LF words. Both a mirror effect and context effects were found in this experiment for accuracy data. However, the lack of any interaction leads one to conclude that the mirror effect, as measured by hit and correct rejection rates, is not affected by changes in environmental context. Another interesting issue brought about by both accuracy and response time data concerned the "rearranged" different context. It was found that a context composed of a novel combination of previously seen context components (screen colours and locations) acts in a similar manner to the same context itself while a completely novel context does not. This, in addition to the estimates of decision criterion data (B"D scores) implies the importance of the types of contexts used and indicates evidence for the use of different response criteria for at least two different types of contexts as the B"D scores in the novel context condition differed remarkably from those in the same-old and rearranged-old context conditions. Lastly, the response time data showed evidence for the mirror effect (main effect of word frequency) in tests for only old items while context effects were found for only new items. This pattern has no obvious explanation and is considered further in Experiment 3.

Experiment 2

Experiment 2 addressed concerns raised in Experiment 1 regarding the "rearranged" context condition. Since a context that consists of a unique combination of context elements previously seen at study acts in a similar manner to the same context that was seen at study, the approach taken next was to vary the level of similarity of the rearranged context. In Experiment 1, two major dimensions were combined to create each context; screen colours and screen location. Experiment 2 manipulated context in such a way as to create four different context conditions at test. The same context condition was again present in addition to the "novel" context condition. However, there were two intermediate context conditions. The first intermediate context condition had the same screen colours as at study with a completely novel screen position (not seen at study). The second intermediate context condition had the same screen position as at study with completely novel screen colours. Using this manipulation, it was possible to assess the relative contributions of the two major dimensions used to create the environmental context. Additionally, this manipulation showed how varying the similarity of the different contexts to the same context condition influences recognition performance using measures of accuracy and response time.

Method

Participants. Sixty-four Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit or payment for participation.

Stimuli and Apparatus. The stimuli and apparatus were the same as used in Experiment 1.

Procedure. The procedure was identical to Experiment 1 except that in the test phase, equal proportions of targets and distractors were tested in four possible contexts. The first context (same context condition) consisted of the same screen colours and location as was used during the learning phase. The second context (old colour condition) consisted of the colours that were used during learning paired with a "novel" screen position (in the centre of the screen). The third context (old location condition) had novel screen colours (white text on black background) paired with the screen location used during learning. The last context (new context condition) had both new screen colours (white text on black background) and location (centre of screen). For distractors, "old" context components meant that the components were used during learning.

Design. A 2 x 4 within-subjects factorial design was used for this experiment. The two factors were word frequency (high vs. low) and type of context (same-old vs. old colour vs. old location vs. novel).

Results and Discussion

Analyses were conducted with mean hit rate (correctly identified targets), mean correct rejection rate (correctly identified distractors), and A' and B"D scores as dependent measures.

Additionally, mean response time (RT) for hits and correct rejections were also analyzed. Means were calculated for each subject for each condition.

Figure 3 displays the proportion of correct responses for new and old tests for LF and HF words in each context condition. As was done in Experiment 1, the proportion of correct responses are arranged in Figure 3 so that the mirror effect is evidenced by a U-shaped pattern of results (averaged across context conditions within frequency groups) for LF and HF words on both "new" and "old" tests. Effects of context condition are shown by higher performance in the new context condition for LF items on "new" tests and higher performance in the same context condition for LF items on "old" tests. A 2 X 4 analysis of variance (ANOVA) based on hit rate showed a highly significant effect of word frequency [$F(1,63) = 41.31$, $MSe = .02$, $p<.001$] and a significant effect of context condition [$F(3,189) = 4.95$, $MSe = .02$, $p=.002$]. The frequency by context condition interaction was not reliable [$F(3,189) = 2.61$, $MSe = .02$, ns]. Paired sample t-tests for hits collapsed across word frequency revealed that performance in the same-old context condition differed significantly from the old colour context condition [$t(127) = 2.093$], the old location context condition [$t(63) = 2.574$], and the novel context condition [$t(127) = 3.598$].

Insert Figure 3 about here

The ANOVA for correct rejection rate also showed significant main effects of word frequency [$F(1,63) = 8.73$, $MSe = .01$, $p=.004$] and of context condition [$F(3,189) = 6.21$, $MSe = .01$, $p<.001$]. As for hits, there was no significant frequency by context condition interaction [$F(3,189) = 1.39$, $MSe = .01$, ns] for correct rejections. Paired sample t-tests for correct rejections collapsed across word frequency showed that performance in the novel context condition differed significantly from the same-old context condition [t(127) = 4.094], the old colour context condition [t(127) = 3.459], and the old location context condition [t(127) = 2.508].

Mean A' values are presented in Table 2 for HF and LF words in the novel, old colour/new location, new colour/old location, and same context conditions. A 2 X 4 ANOVA was conducted for both of the values as dependent measures. Analysis of A' showed a significant main effect of frequency [$F(1,63) = 38.58$, MSe = .01, $p < .001$]. The main effect of context condition [$F(3,189) < 1$, MSe = .01] and the frequency by context condition interaction [$F(3,189) < 1$, MSe = .01] did not approach significance. d' values were also calculated and their pattern was the same as that of the A' scores.

Insert Table 2 about here

Mean B"D values are presented in Table 3 for HF and LF words in the novel, old colour/new location, new colour/old location, and same context conditions. The analysis of B"D (the estimate of the decision criterion associated with A') revealed significant effects of frequency [$F(1,63) = 4.04$, MSe = .23, $p < .05$], and context condition [$F(3,189) = 7.74$, MSe = .20, $p < .001$], and a significant frequency by context interaction [$F(3,189) = 3.48$, MSe = .23, $p = .017$]. Paired sample t-tests for B"D values revealed that for HF words, the only reliable difference was between the old colour/new location condition and the novel context condition [t(63) = 1.68]. However, for LF words, the same old context condition was significantly different from the old colour/new location condition [t(63) = 2.68], the new colour/old location condition [t(63) = 2.24], and the novel context condition [t(63) = 4.72]. Additionally, both the old colour/new location condition [t(63) = 2.38] and the new colour/old location condition [t(63) = 2.21] were reliably different from the novel context condition.

Insert Table 3 about here

The mean RTs for correct new and old responses in each condition are displayed in Figure 4. A 2 X 4 ANOVA was performed separately for hits and correct rejections with mean RTs for hits as the dependent measure and with mean RTs for correct rejections as the dependent measure. The mean response times are arranged in Figure 4 so that the mirror effect is evidenced by an inverted U-shaped pattern. The analysis of the mean RTs for hits revealed a main effect of frequency [$F(1,63) = 15.09$, $MSe = 111{,}744.2$, $p<.001$] and a main effect of context condition [$F(3,189) = 17.05$, $MSe = 89{,}111.7$, $p<.001$]. The frequency by context condition interaction was not reliable [$F(3,189) = 1.52$, $MSe = 108{,}707.5$, ns]. Paired sample t-tests for hits collapsed across word frequency revealed that RTs were significantly different between the same-old context condition and the old colour context condition [$t(127) = 4.511$], the old location context condition [$t(127) = 3.589$], and the novel context condition [$t(127) = 6.807$]. Additionally, the mean RT for the old location context condition was found to be significantly greater than the novel context condition [$t(127) = 3.773$]. These results can be seen in Figure 4 as a consistent pattern is displayed for hits with the same old context condition having the longest RT followed by the new colour/old location condition and the old colour/new location condition, respectively. The new context condition consistently had the shortest RT. The pattern of RTs for hits was different than that found in Experiment 1. Notably, the RTs of items presented in an old location were faster than those presented in a new location whereas in Experiment 1 there was no difference between items presented in old and new locations. The eye movements required by participants to locate an item presented in an old location may be a significant factor influencing RTs.

Insert Figure 4 about here

The analysis of the mean RTs for correct rejections revealed a significant main effect of frequency [$F(1,63) = 9.14$, MSe = 85,592, $p=.004$] and a significant main effect of context condition [$F(3,189) = 22.38$, MSe = 66,932.2, $p<.001$]. The frequency by context condition interaction was not significant [$F(3,189)<1$, MSe = 74,992.6]. Paired sample t-tests for correct rejections collapsed across frequency showed that RTs in the same-old context condition were significantly different from those in the old colour context condition [$t(127) = 6.495$], the old location context condition [$t(127) = 3.822$], and the novel context condition [$t(127) = 6.072$]. Also, the old location context condition was found to be significantly different from the old colour context condition [$t(127) = 3.265$] and the novel context condition [$t(127) = 3.767$]. The pattern of RTs for correct rejections was similar to that of hits and also to the pattern of correct rejections in Experiment 1 in that words presented in old locations had a slower RT than those presented in a new location. Overall, the presence of a mirror effect is consistent with the recognition tests of old items but not with the tests of new items in Experiment 1. Also, the presence of context effects in tests of both new and old items is consistent with the recognition tests of new items but not old items in Experiment 1.

Experiment 2 separated the two major aspects contributing to environmental context (screen colour and location) to create intermediate contexts that differed in their similarity to the same context condition at test. Similar to Experiment 1, both a mirror effect and context effects were found for accuracy data. The addition of the partial context conditions (i.e. old colour condition and old location condition) led to an interesting pattern of results. On tests of new items (i.e. correct rejections), the partial context conditions were treated like the same-old context condition which suggests that participants were more likely to identify new words only when they were accompanied by a completely novel context. For tests of old items (i.e. hits) the partial context conditions were treated in a similar manner as the novel context condition, suggesting that participants were more likely to identify old words only when accompanied by a same-old context. Thus, it seems that only a completely novel context provides an advantage for new test items while only a completely old-looking context provides an advantage for old test items (it needs not be the same exact context as at study; see the proportion correct data of Experiment 1). The partial context conditions appear to exert little influence, if any, upon recognition decisions.

The pattern of A' scores in Experiment 2 was similar to that found in Experiment 1, which again shows a greater sensitivity in the recognition of LF words versus HF words. The B''_D scores observed in Experiment 2 suggest criterion changes are occurring across the different conditions (the same was indicated in Experiment 1 as in both experiments, the novel context condition produced the highest B''_D values in both LF and HF conditions). However, the pattern differed from Experiment 1 with respect to LF and HF same-old context conditions.

Another issue of interest in Experiment 2 was response times. One major finding was the occurrence of a mirror effect that was not observed in Experiment 1. Also, in the test conditions that contained old screen locations, response time was slower than the context conditions that had a centre-screen location (novel and old colour conditions). This pattern of results implies that the eye movements involved in locating the test probe serve as a major contribution to response time scores. With this in mind, if participants focus on the centre of the screen between test probes, they would locate test probes in the centre of the screen significantly faster than test probes presented in one of the possible four corners of the screen. Since it is difficult to determine a true pattern of response times across the different context conditions when eye movements could be creating a major confound, Experiment 3 eliminated the use of different screen locations during the recognition test in order to keep eye movements at a minimum.

Experiment 3A and 3B

Experiment 3 served as a replication of Experiment 1 with some modifications that attempted to clarify issues raised through Experiments 1 and 2. Due to the pattern of response times in Experiment 2 that suggested eye movements significantly affect the observed response times, location was left out as a variable context component. This was done to eliminate the eye movements necessary to locate words that were presented in various corners of the computer screen in Experiments 1 and 2 and thus attempted to provide more accurate and interpretable response time data. It was expected that, once eye movements were minimized, the pattern of response times would be similar to that of the proportion correct data. That is, LF words would be recognized faster than HF words, and new and old words would be recognized faster in novel and same-old contexts, respectively.

Also, Experiment 3 had only two test contexts; same-old and novel. In Experiment 1, participants treated the rearranged context in the same way that they treated the same old context. Since participants do not associate the TBR words with specific contexts but rather the overall appearance of the contexts, it was not necessary to again include the rearranged context.

Method

Participants. Forty-four Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool to participate in Experiment 3A and forty were recruited for Experiment 3B. Students earned course credit for participation.

Stimuli and Apparatus. The stimuli and apparatus were the same as used in Experiment 1 except that context elements consisted of the four different possibilities for foreground colour and background colour only. All stimuli were presented in the centre of the computer screen.

Procedure. The procedure for Experiment 3A was identical to Experiment 1 except that context in Experiment 3A was defined as the unique combination of only foreground and background colour. Since this allowed only 16 different context combinations, 12 different contexts were used twice to allow for 24 word pair presentations during the learning phase. In Experiment 1 it was demonstrated with the rearranged context that participants do not associate presented words with specific contexts but rather the old or new appearance of the contexts they are presented in. Therefore it should not make a difference whether a learning context is presented with 1 or 2 word pairs.

Experiment 3B was identical to Experiment 3B with one exception. The duration of the word pair presentations in Experiment 3B was reduced from 3 seconds to 2 seconds. The reason for this change was the high level of performance found in Experiment 3A. To avoid any ceiling effects, viewing time for each word pair during the learning phase was shortened. For Experiments 3A and 3B, the written instructions given to participants were modified so as not to mention varying locations.

Design. A 2 x 2 within-subjects factorial design was used for this experiment. The two factors were word frequency (high vs. low) and type of context (same-old vs. novel).

Results and Discussion

Experiment 3A. Figure 5 displays the proportion of correct responses for new and old tests for LF and HF words in both context conditions. As in Experiment 1and 2, the proportion of correct responses are arranged in Figure 5 so that the mirror effect is evidenced by a U-shaped pattern of results for LF and HF words on both "new" and "old" tests. Effects of context condition are shown by higher performance in the novel context condition on "new" tests and lower performance in the novel context condition on "old" tests. A 2 X 2 analysis of variance (ANOVA) based on hit rate showed a significant effect of word frequency [$F(1,43) = 26.6$, MSe = .007, $p<.001$] and a significant effect of context condition [$F(1,43) = 14.4$, MSe = .008, $p<.001$]. However, there was no significant frequency by context condition interaction [$F(1,43) = 2.97$, MSe = .005, ns]. The ANOVA for correct rejection rate also showed a significant effect of word frequency [$F(1,43) = 11.8$, MSe = .008, $p=.001$] and a significant effect of context condition [$F(1,43) = 36.1$, MSe = .011, $p<.001$]. Similar to hits, there was no significant frequency by context condition interaction [$F(1,43) = 3.66$, MSe = .006, ns] for correct rejections.

Insert Figure 5 about here

Mean A' values are presented in Table 4 for HF and LF words in the novel and the same-old context conditions. A 2 X 2 ANOVA analysis of A' showed a significant effect of frequency [$F(1,43) = 32.14$, MSe = .002, $p<.001$] and a main effect of context condition [$F(1,43) = 6.53$, MSe = .002, $p = .014$] while the frequency by context condition interaction [$F(1,43)<1$, MSe = .002] did not approach significance.

Insert Table 4 about here

Mean B"D (the estimate of the decision criterion associated with A') values are presented in Table 5 for HF and LF words in the novel and the same-old context conditions. A 2 X 2 ANOVA analysis of B"D showed a significant effect of context condition [$F(1,43) = 50.11$, MSe = .14, $p<.001$]. The main effect of frequency [$F(1,43)<1$, MSe = .20] and the frequency by context condition interaction for B"D [$F(1,43) = 3.25$, MSe = .11, ns] did not reach significance.

Insert Table 5 about here

The mean RTs for correct new and old responses in each condition are displayed in Figure 6. A 2 X 2 ANOVA was performed separately for hits and correct rejections. The analysis of the mean RTs for hits revealed a significant main effect of frequency [$F(1,43) = 21.53$, MSe = 28,657.6, $p<.001$] but no main effect of context condition [$F(1,43) = 3.51$, MSe = 26,671.9, ns]. However, a frequency by context condition interaction was found [$F(1, 43) = 8.87$, MSe = 25, 670.8, $p = .005$]. Paired sample t-tests for these RT values revealed that for LF words, the same-old context condition was not significantly faster than the novel context condition [$t(43) = .90$]. For HF words, the novel context condition was significantly faster than the same-old context condition [$t(43) = -2.99$] showing a significant difference opposite from the predicted direction. These results are seen in Figure 6 as a faster RT for LF targets than HF targets and a difference between the context conditions only within the HF condition.

Insert Figure 6 about here

The analysis of the mean RTs for correct rejections revealed a significant effect of context condition [F(1,43) = 22.04, MSe = 44,927.2, p<.001] and an effect of frequency [F(1,43) = 15.19, MSe = 31,537.6, p<.001], but no frequency by context condition interaction [F(1,43) = 1.67, MSe = 27,498.6, ns]. These results are evidenced by the faster RT for "new" items presented in a novel context compared to the same-old items for both LF and HF words, and faster RTs for LF versus HF words.

Experiment 3B. Figure 7 displays the proportion of correct responses for new and old tests for LF and HF words in both context conditions. A 2 X 2 analysis of variance (ANOVA) based on hit rate showed a significant effect of word frequency [F(1,39) = 27.41, MSe = .008, p<.001] and a significant effect of context condition [F(1,39) = 8.56, MSe = .011, p = .006]. However, there was no significant frequency by context condition interaction [F(1,39)<1, MSe = .005]. The ANOVA for correct rejection rate also showed a significant effect of word frequency [F(1,39) = 22.43, MSe = .009, p<.001] and a significant effect of context condition [F(1,39) = 18.34, MSe = .011, p<.001]. Similar to hits, there was no significant frequency by context condition interaction [F(1,39)<1, MSe = .006] for correct rejections.

Insert Figure 7 about here

Mean A' values are presented in Table 4 for HF and LF words in the novel and the same-old context conditions. A 2 X 2 ANOVA was conducted for these values. Analysis of A' showed a significant effect of frequency [F(1,39) = 45.50, MSe = .003, p<.001] while the main effect of context condition [$F(1,39) = 2.54$, $MSe = .004$, ns] and the frequency by context condition interaction [$F(1,39) = 2.44$, $MSe = .004$, ns] did not approach significance.

Mean B''_D values are presented in Table 5 for HF and LF words in the novel and the same-old context conditions. Similar to Experiment 3A, the analysis of B''_D showed a significant effect of context condition [$F(1,39) = 21.9$, $MSe = .16$, $p<.001$] while the main effect of frequency [$F(1,39)<1$, $MSe = .12$] and the frequency by context condition interaction for B''_D [$F(1,39)<1$, $MSe = .10$] did not reach significance.

The mean RTs for correct new and old responses in each condition are displayed in Figure 8. A 2 X 2 ANOVA was performed separately for hits and correct rejections. The analysis for hits revealed a significant main effect of frequency [$F(1,39) = 4.79$, $MSe = 87{,}806.9$, $p = .035$], but no main effect of context condition [$F(1,39)<1$, $MSe = 57{,}994.5$] and no frequency by context condition interaction [$F(1,39) = 1.68$, $MSe = 39{,}788.9$, ns] were found. These results are seen in Figure 8 as a faster RT for LF targets than HF targets and no significant differences between the novel and same-old context conditions.

Insert Figure 8 about here

Identical to Experiment 3A, the analysis of the mean RTs for correct rejections revealed a significant effect of context condition [$F(1,39) = 16.62$, $MSe = 46{,}097.7$, $p<.001$] and an effect of frequency [$F(1,39) = 13.88$, $MSe = 29{,}851.3$, $p = .001$], but no frequency by context condition interaction [$F(1,39)<1$, $MSe = 30{,}068.4$]. These results are evidenced by the faster RT for "new" items presented in a novel context compared to the same-old items for both LF and HF words, and faster RTs for LF versus HF words.

Experiments 3A and 3B served to elucidate and replicate findings previously obtained in Experiments 1 and 2. Using a simpler design that had only two context conditions, same-old and new, Experiments 3A and 3B showed a stable pattern of hits and correct rejections. Similar to Experiment 1, both the mirror effect and effects of context were observed with no interaction. This provides further support for the view that these are two separate processes influences on recognition memory. Even though Experiment 3B's shorter exposure times during learning attenuated the high performance levels found in Experiment 3A, the pattern of hits and correct rejections in Experiment 3A was replicated.

A' values in Experiments 3A and 3B showed the same effect of frequency as observed in Experiments 1 and 2, again indicating participants' greater sensitivity to LF as opposed to HF words. Experiment 3A also showed an effect of context condition, however, when the exposure time was reduced to 2 sec per word pair in Experiment 3B, this effect was not observed. Thus, the stable finding is participants' increased sensitivity to LF words.

The B''_D values obtained in Experiments 3A and 3B provide further support for the use of different decision criteria by participants when there is a change in context at the recognition test. Both Experiments 3A and 3B produced a highly significant effect of context, indicating more conservative responses in the novel context condition as opposed to the same-old context condition. Experiments 1 and 2 also showed significant effects of context suggesting that different response criteria are used depending upon the nature of the context at test.

The response time data in Experiment 2 suggested that eye movements (length of time participants took to locate words presented in various areas of the screen) may have been playing a role in Experiments 1 and 2. Therefore, screen location was dropped as a context element for Experiments 3A and 3B. It was believed that this would provide a cleaner pattern of response times as eye movements would not be as significant a factor. For tests of novel items, there was both an effect of frequency and an effect of context. That is, novel items were recognized faster when presented in a novel context and when they were LF words. The results for tests of old items were less clear. The only consistent effect over Experiments 3A and 3B was an effect of frequency (LF words were recognized faster than HF words). Thus, while word frequency produced consistent effects on response times that mirrored the effects of accuracy, context only had a consistent effect on response times for new items.

One issue that has not been approached as yet in the present experiments is that of confidence judgments. It is clear now that participants use different response criteria depending on the type of context old and new items are presented in. What is also clear is the pattern of hits and correct rejections on the recognition tests employed in these experiments. However, how participants would rate their confidence on these recognition decisions is unknown. Experiment 4 was designed to address this issue.

Experiment 4

Experiment 4 was a replication of Experiment 3B. Experiment 4 differed in that participants were provided with six response alternatives during the recognition test to allow for a comparison between level of confidence and new/old responses. The response options consisted of sure-new, maybe-new, guess-new, guess-old, maybe-old, and sure-old so participants can indicate the appropriate level of confidence for each recognition decision. Since the judgments participants were asked for are more complex than those used in the previous experiments, response time was not a focus of Experiment 4.

It was difficult to predict how participants will rate their confidence. There could very well be no relationship between confidence and the pattern of results we have observed thus far. However, one can speculate about possible relationships. If confidence is a memory-based phenomenon and reflects the participants sensitivity to different items, level of confidence would be expected to correlate with the pattern of A' scores that consistently show a greater sensitivity for LF as opposed to HF items (i.e. confidence may be greater for LF items than HF items but unaffected by context). On the other hand, if confidence judgments are influenced primarily by response bias, they may reflect where participants place their response criteria and would be expected to follow the pattern of results shown by B''_D scores which consistently show more conservative responding for test items presented in a novel context (i.e. participants may be more confident when responding conservatively). Finally, confidence judgments may be found to be an artifact of proportion correct data. If this is the case, then it would be expected that confidence would be higher for LF items compared to HF items. Additionally, novel context items would show greater confidence on tests of new items while same-old context items would have higher confidence ratings for test of old items.

Method

Participants. Forty-four Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit for participation.

Stimuli and Apparatus. The stimuli and apparatus were the same as used in Experiment 3B except that 6 keys (instead of 2) were exposed to the participants to indicate the 6 possible response options.

Procedure. The procedure was identical to Experiment 3B except that when participants were prompted for their old/new responses, they also indicated their level of confidence. Three keys (R, T, and Y) were designated for "old" responses. These represented a sure-old (participants are sure the item is old), a maybe-old (participants think that an item is old but aren't sure), and a guess-old (participants really don't know but are saying "old") response. Three keys (U, I, and O) were designated for "new" responses. These represented a sure-new (participants are sure the item is new), a maybe-new (participants think that as item is new but aren't sure), and a guess-new (participants really don't know but are saying "new") response. The written instructions given to participants were modified to explain the confidence judgments they were asked to make.

Design. A 2 x 2 within-subjects factorial design was used. The two factors were word frequency (high vs. low) and type of context (same-old vs. novel).

Results and Discussion

Analyses were conducted with mean hit rate and mean correct rejection rate as dependent measures. Also, the confidence level attributed to hits and correct rejections were also analyzed. Means were calculated for each subject in each condition.

Figure 9 shows the overall proportion of correct responses for new and old tests for LF and HF words in the same-old and novel context conditions collapsed over confidence judgments. Effects of context can be seen as higher performance in the novel context condition on "new" tests (tests of novel items) and lower performance in the novel context condition on "old" tests (tests of old items). A 2 X 2 ANOVA based on hit rate showed a significant effect of word frequency [$F(1,43) = 38.88$, $MSe = .005$, $p<.001$] and a significant effect of context condition [$F(1,43) = 10.81$, $MSe = .012$, $p=.002$]. The frequency by context condition interaction was not significant [F(1,43)<1, MSe = .006]. Similar to hits, the ANOVA for correct rejection rate also showed a significant effect of word frequency [F(1,43) = 38.05, MSe = .005, p<.001] and a significant effect of context condition [F(1,43) = 19.80, MSe = .013, p<.001] but no significant frequency by context condition interaction [F(1,43) <1, MSe = .006].

Insert Figure 9 about here

The mean confidence ratings for new and old responses in each condition are displayed in Table 6. A value of "3" was assigned to all "sure" responses, a value of "2" to all "maybe" responses, and a value of "1" to all "guess" responses. A 2 X 2 ANOVA was performed separately for hits and correct rejections. The analysis of the mean confidence ratings for hits revealed a significant effect of frequency [F(1,43) = 26.33, MSe = .027, p<.001] and a marginally significant effect of context condition [F(1,43) = 4.14, MSe = .036, p=.048]. No frequency by context condition interaction was found for mean confidence ratings of hits [F(1,43) <1, MSe = .011]. Similar to hits, the analysis of the mean confidence ratings for correct rejections showed a significant effect of frequency [F(1,43) = 30.86, MSe = .037, p<.001]. However, the main effect of context condition only approached significance [F(1,43) = 3.029, MSe = .031, p=.089]. Again, the frequency by context condition interaction did not approach significance [F(1,43) < 1, MSe = .023].

Insert Table 6 about here

The analyses of the mean confidence ratings for hits and correct rejections showed a strong effect of word frequency (LF items having a higher mean confidence level than HF items) and a pattern of results across context conditions consistent with the proportion correct data (higher mean confidence for old items in the same-old context over old items presented in a novel context and higher mean confidence for new items in a novel context over new items in an old context). This prompted the question of whether the effects of context condition and frequency observed in the proportion correct data were a result of LF words having more high-confidence responses than HF words and context-test probe consistent items (e.g. old word presented in the same-old context; new word presented in a novel context) having more high-confidence responses than context-test probe inconsistent items (e.g. old word presented in a novel context; new word presented in an old context). To answer this question, the proportion correct data was re-analyzed conditionalized upon high-confidence responses. In other words, only responses given a high-confidence rating were included, equating all conditions on confidence level.

Figure 10 shows the proportion of high-confidence correct responses for new and old tests for LF and HF words in the same-old and novel context conditions. The mirror effect is evidenced by a U-shaped pattern of results for LF and HF words on both "new" and "old" tests. Effects of context are also found as performance is higher in the novel context condition on "new" tests (tests of novel items) and lower performance in the novel context condition on "old" tests (tests of old items). A 2 X 2 ANOVA based on hit rate showed a significant effect of word frequency [$F(1,43) = 63.82$, $MSe = .010$, $p<.001$] and a significant effect of context condition [$F(1,43) = 10.03$, $MSe = .016$, $p=.003$]. The frequency by context condition interaction was not found to be significant [$F(1,43) <1$, $MSe = .008$]. Similar to hits, the ANOVA for correct rejection rate also showed a significant effect of word frequency [$F(1,43) = 48.23$, $MSe = .014$, $p<.001$] and a significant effect of context condition [$F(1,43) = 14.02$, $MSe = .014$, $p=.001$] but no significant frequency by context condition interaction [$F(1,43) <1$, $MSe = .008$].

Insert Figure 10 about here

Considering the proportion correct data, Experiment 4 provided further support for the notion of the mirror effect and effects of context being two separate influences in recognition memory. Once again, both the mirror effect and effects of context were observed with no interaction. However, Experiment 4 also required that participants make decisions as to their level of confidence in their recognition decisions. Analysis of the mean confidence ratings across conditions revealed a highly significant effect of word frequency and a pattern of differences between context conditions consistent with the proportion correct data of this and the previous experiments. However, the word frequency manipulation seemed to have a greater effect on confidence judgments than the context manipulations. While this is consistent with the suggestion that confidence judgments are memory-based (based on differences in sensitivity), the small influence of context suggests that they may be partially a result of response bias as well.

While it was suggested that effects of frequency and context observed in the proportion correct data may be a result of the differences in the number of high-confident judgments across conditions (i.e. more high-confident responses for LF compared to HF items and more high confident responses for context-test probe consistent conditions compared to context-test probe inconsistent items), an analysis of the proportion correct data using only high-confident responses showed that the mirror effect and effects of context are observed even when all conditions are equated for level of confidence.

Experiments 3 and 4 demonstrate the stability of the two effects of interest in the present study. The influence of word frequency is quite apparent in considering Experiments 1 through 4.

However, the effects of context are less clear when different types of contexts are added to the standard same-old and novel context conditions (i.e. the multiple contexts used in Experiments 1 and 2). One question that arises is whether increasing the number of times a particular context is seen during study relative to other old contexts will increase the effects of changing context from study to test for items presented in that context. Experiment 5 was designed to address this question.

Experiment 5

Experiment 1 showed that a context composed of a novel combination of previously seen context elements is treated in a similar manner to the same-old context. This suggests that the importance of a context lies in whether it appears old or whether it appears novel to participants. Participants in the present study have employed different response criteria when presented with items in same-old and different context conditions. If the appearance of a context as "new-looking" or "old-looking" is indeed what influences recognition decisions in differing contexts, one can speculate as to what makes a context appear old. We have demonstrated that using certain screen colours at study followed by those same screen colours as the same-old context condition and colours never seen during study as the novel context condition is effective at producing effects of context. An attempt to create two levels of strength for same-old context conditions (i.e. of two different types of same-old contexts, one would seem "more old" than the other) was made in Experiment 5. The first was contexts presented only once during study (as was used in the earlier experiments) while the second was contexts presented eight times during study. The question of interest is whether presenting a particular context multiple times during study will increase the effect of changing that context at test over contexts that are presented only once during study.

Murnane and Phelps (1995) manipulated item strength by varying the number of times items were presented in their respective contexts during the study phase in one experiment. Participants would witness spaced repetitions of items in the study list from one to five times. They found that as the number of repetitions increased for an item and its context, the magnitude of the context effect at test increased. They interpreted these results as supportive for global activation models of recognition memory. However, Murnane and Phelps repeated word pairs and contexts together.

Experiment 5 attempts to determine whether multiple presentations of a context (without a specific word pair being presented with it) will influence recognition decisions at test without items being presented multiple times. In this experiment, items were presented in one of two types of contexts. For each study trial, one context was presented eight times with different word pairs (8x context condition) and eight other contexts were presented only once with different word pairs (1x context conditions). Thus, at test, targets and distractors were presented in either the same-old context (either the 1x or the 8x condition) or a novel context. The question of interest was whether items originally presented in a context that was repeated at study are more affected by a change in context at test than items presented in a context presented only once at study. If the repetition of a study context can influence recognition decisions, then the results of Experiment 5 should reveal a similar pattern to that of Murnane and Phelps (i.e. the magnitude of the context effect increases with increased repetitions). Indeed, global matching models would posit that a context repeated eight times would lead to greater familiarity at test than a context presented only once. However, if it was the repetition of items with the contexts that led to the higher recognition performance in the Murnane and Phelps (1995) study then there should be no difference between the single and the repeated context conditions.

Method

Participants. Forty Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit for participation.

Stimuli and Apparatus. The stimuli and apparatus were identical to that used in Experiment 3B.

Procedure. The procedure was identical to Experiment 3B except that, during each study phase, one context was presented with 8 different word pairs. The remaining 8 tested word pairs (not including primacy and recency buffers) were each presented in their own individual contexts. Additionally, there were only three primacy and three recency buffers used (as opposed to the four used in the previous experiments), each presented in their own context. Thus, each study phase had one 8x context and fourteen 1x contexts (however, only eight were tested). Therefore, at test LF and HF old items could fall into one of four different categories; 1x study context presented in the same context condition, 1x study context presented in a novel context, 8x study context presented in the same context condition, and 8x study context presented in a novel context condition. HF and LF new items, on the other hand, could fall into one of three categories; new items presented in a novel context, new items presented in a 1x old context, and new items presented in a 8x context condition.

Design. Tests of old items were analyzed as a 2 x 2 x 2 within-subjects factorial design with the three factors word frequency (high vs. low), type of test context (same-old vs. novel), and number of study context repetitions (1x vs. 8x). Tests of new items were analyzed as a 2 x 3 within-subjects factorial design with the two factors word frequency (high vs. low) and type of test context (novel vs. 1x old vs. 8x old).

Results and Discussion

Analyses were conducted with mean hit rate and mean correct rejection rate as dependent measures. Additionally , mean response time (RT) for hits and correct rejections were also analyzed. Means were calculated for each subject in each condition.

Figure 11 shows the proportion of correct responses for tests of old LF and HF words (targets) in both test context conditions. Half of these words had been presented in a 1x study context condition and half had been presented in a 8x study context condition. Effects of word frequency are evidenced by greater performance in the LF word condition over the HF word condition for both 1x study context items and 8x study context items. Effects of context condition are shown by greater performance for all types of old items when presented in the same-old context condition at test compared to the novel context condition. A 2 X 2 X 2 ANOVA based on hit rate showed a significant effect of word frequency [$F(1,39) = 54.91$, $MSe = .008$, $p<.001$] and a significant effect of test context condition [$F(1,39) = 7.546$, $MSe = .019$, $p=.009$]. There was no effect of study context repetitions [$F(1,39) = 1.186$, $MSe = .013$, ns]. All interactions were found to be non-significant using $\alpha = .05$.

Insert Figure 11 about here

The proportion of correct responses for LF and HF new items (distractors) in the three test context conditions are presented in Figure 12. An effect of word frequency is demonstrated by the greater performance of LF words over HF words. A context effect is also evident as the proportion of correctly identified distractors in the novel context condition exceeds that in both the 1x and the 8x context conditions. A 2 X 3 ANOVA for correct rejection rate showed a significant effect of word frequency [$F(1,39) = 8.966$, $MSe = .012$, $p=.005$] and a significant effect of context condition [$F(2,78) = 4.165$, $MSe = .007$, $p=.019$]. The frequency by context condition interaction did not approach significance [$F(2,78)<1$, $MSe = .009$] for correct rejections. Paired same t-tests collapsed across context condition revealed that the novel context condition differed significantly from both the 1x old context condition [$t(79) = 2.609$] and the 8x context condition [$t(79) = 2.286$]. However, the 1x and 8x context conditions did not differ [$t(79) = .263$].

Insert Figure 12 about here

The mean RTs for correct responses for tests of old LF and HF words (targets) in both test context conditions are displayed in Figure 13. Like the proportion correct data, half of these words had been presented in a 1x study context condition and half had been presented in a 8x study context condition. The 2 X 2 X 2 ANOVA for the mean RTs for hits revealed a significant effect of word frequency [$F(1,39) = 33.11$, $MSe = 73,637.1$, $p<.001$] evidenced by faster performance in the LF conditions compared to the HF conditions. The main effects of context condition [$F(1,39) = 1.217$, $MSe = 46,849.7$, ns] and study context repetition [$F(1,39) = 3.939$, $MSe = 52,593.9$, ns] were both non-significant. A frequency by study context repetition interaction was found [$F(1,39) = 10.22$, $MSe = 42,198.2$, $p=.003$] indicating a greater difference in LF and HF response times for 1x study context words compared to 8x study context words. Pairwise t-tests confirmed these findings as participants took significantly longer to respond to HF words as opposed to LF words for both same-old [t(39) = 4.96] and novel [t(39) = 3.60] context conditions for the 1x study items while RT was only significantly greater for HF compared to LF words in the same-old context condition [t(39) = 3.09] for the 8x study items. This difference was not significant for novel context 8x study items [t(39) = 1.57]. Additionally, there was a significant study context repetition by test context condition interaction [F(1,39) = 25.53, MSe = 40,034.3, p<.001] evidenced by a faster response time in the same-old test condition for words presented in the 8x study context condition but a faster response time in the novel context test condition for words presented in the 1x study context condition. Pairwise t-tests confirmed this by showing faster RTs for same-old context LF [t(39) = 2.23] and HF [t(39) = 2.12] items compared to novel context items for the 8x study condition. This pattern was reversed for the 1x study items as RTs were faster for novel context LF [t(39) = 2.16] and HF [t(39) = 3.01] items compared to same-old context items. The frequency by test context condition interaction [F(1,39) = 2.52, MSe = 42,498.9, ns] and the frequency by study context repetition by test context condition interaction [F(1,39)<1, MSe = 49,448.5] did not reach significance.

Insert Figure 13 about here

The mean response time for LF and HF new items (distractors) in the three test context conditions are presented in Figure 14. A context effect is evident as the response time for correctly identified distractors in the novel context condition is faster than that in both the 1x and the 8x context conditions. However, there was only a small difference in response times between LF and HF distractors. A 2 X 3 ANOVA for correct rejection response time showed a significant effect of context condition [F(2,78) = 7.95, MSe = 150,124.4, p=.001]. However the main effect of word frequency [$F(1,39) = 1.29$, $MSe = 130,168.3$, ns] and the frequency by context condition interaction [$F(2,78)<1$, $MSe = 95,805.1$] did not approach significance for correct rejections. Paired sample t-tests collapsed across context condition revealed that the novel context condition differed significantly from both the 1x old context condition [$t(79) = 3.840$] and the 8x context condition [$t(79) = 4.531$]. However, 1x and 8x context conditions did not differ [$t(79) = .59$].

Insert Figure 14 about here

Experiment 5 again demonstrates the statistical independence of the mirror effect and effects of environmental context in terms of hit rate and correct rejection rate. Both the mirror effect and effects of context were present with no interaction. Additionally, the number of times contexts were seen at study (some contexts were seen eight times and others only once) was manipulated in Experiment 5. For hits no effect of repetition was found, indicating that multiple repetitions of a particular context during study do not increase the effect of changing context. There was also no difference between the correct rejection rates of novel items presented in a 1x context and items presented in an 8x context. The differences in the results observed here and those found by Murnane and Phelps may be due to the fact that Murnane and Phelps used multiple presentations of the words and their context during study, perhaps strengthening the word-context association. On the other hand, it may simply have been the repetition of the words that led to a greater effect of context in their study. In any case, a repeated context on its own does not seem to differently affect recognition decisions more than a context seen only once during study.

Response time was also analyzed in Experiment 5 in the hopes of clarifying the results found in Experiments 3A and 3B. While Experiments 3A and 3B showed both an effect of context and frequency on tests of novel items, Experiment 5 showed only an effect of context. The response time results of tests of old items showed an effect of frequency (as was found in both Experiments 3A and 3B) and context repetition interacted with both frequency and context. Therefore, the consistency of context having an effect on response time for correct rejections and frequency having an effect on response time for hits suggest that context has a greater effect on tests of new items while frequency has a greater effect on tests of old items in terms of response time. The reason for this, while interesting to note, is unclear at this point.

Experiments 1 through 5 have used the stimulus manipulation of word frequency to demonstrate the mirror effect. However, it is by no means the only manner by which the mirror effect has been found. Another manipulation that has been used (e.g. Glanzer & Adams, 1985; Hockley, 1994) is nouns versus nonnouns. This stimulus manipulation was used in Experiment 6 to determine the generalizability of the current findings.

Experiment 6

Glanzer & Adams (1985) showed that the mirror effect can be found for a variety of stimulus manipulations. The purpose of Experiment 6 was to show that the independence of context effects and the mirror effect shown for word frequency will generalize to another manipulation of the mirror effect. In Experiment 6, recognition performance for nouns and nonnouns was compared. This stimulus manipulation has been shown to produce the mirror effect in measures of both accuracy and response latency (Hockley, 1994).

Method

Participants. Forty-six Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit for participation.

Stimuli and Apparatus. With the exception of the type of words used, the stimuli and apparatus were identical to that used in Experiment 3B. The words for Experiment 6 were selected from a 1000-word version of the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982) consisting of common two-syllable words evenly divided between nouns and nonnouns. Nonnouns consisted of adjectives, verbs, and adverbs. According to the norms of Friendly et al., which rate words on scales from 1 to 7, the nouns have a higher mean value of imagery (4.93 vs. 3.56) and concreteness (5.09 vs. 3.49) than do nonnouns. The mean word length was similar for nouns (6.11 letters) and nonnouns (6.07). According to the Kucera and Francis (1967) counts, the nonnouns have a higher mean natural language frequency (86.19 per million) than nouns (62.63).

Procedure. The procedure was identical to Experiment 3B except for the change in the stimulus manipulation.

Design. A 2 x 2 within-subjects factorial design was used for this experiment. The two factors were word type (noun vs. nonnoun) and type of context (same-old vs. novel).

Results and Discussion

Analyses were conducted with mean hit rate, mean correct rejection rate, and A' and B''_D scores as dependent measures. Additionally, mean response time (RT) for hits and correct rejections were also analyzed. Means were calculated for each subject in each condition.

Figure 15 shows the proportion of correct responses for new and old tests for nouns and nonnouns in both context conditions. Effects of environmental context are again shown by higher performance in the novel context condition on "new" tests and lower performance in the novel context condition on "old" tests. A 2 X 2 ANOVA based on hit rate revealed a significant effect of word type [$F(1,45) = 8.029$, $MSe = .007$, $p=.007$] and a significant effect of context condition [$F(1,45) = 26.667$, $MSe = .009$, $p<.001$]. The word type by context condition interaction did not approach significance [$F(1,45)<1$, $MSe = .008$]. The ANOVA for correct rejection rate also showed a significant effect of word type [$F(1,45) = 10.956$, $MSe = .007$, $p=.002$] and a significant effect of context condition [$F(1,45) = 18.845$, $MSe = .011$, $p<.001$]. Similar to hits, there was no significant word type by context condition interaction [$F(1,45)<1$, $MSe = .006$] for correct rejections.

Insert Figure 15 about here

Mean A' values are presented in Table 7 for nouns and nonnouns in the novel and same-old context conditions. A 2 X 2 ANOVA of A' revealed a significant effect of word type [$F(1,45) = 12.755$, $MSe = .003$, $p=.001$] while the main effect of context condition [$F(1,45)<1$, $MSe = .002$] and the word type by context condition interaction [$F(1,45)<1$, $MSe = .003$] did not approach significance.

Insert Table 7 about here

Mean B''_D values are presented in Table 7 for nouns and nonnouns in the novel and same-old context conditions. A 2 X 2 ANOVA of B''_D showed a significant effect of context condition

[F(1,45) = 31.820, MSe = .155, p<.001]. The main effect of word type [F(1,45)<1, MSe = .077] and the word type by context condition interaction [F(1,45)<1, MSe = .084] did not approach significance.

The mean RTs for correct new and old responses in each condition are displayed in Figure 16. A 2 X 2 ANOVA was performed separately for hits and correct rejections. The analysis of the mean RTs for hits revealed no effect of word type [F(1,45)<1, MSe = 24,499.6] and no effect of context condition [F(1,45)<1, MSe = 19,191.9]. However, the word type by context condition interaction was found to be significant [F(1,45) = 8.070, MSe = 28,146.9, p=.007]. Paired sample t-tests for these RT values showed that this interaction was a result of a faster RT for old nonnouns when presented in the same-old context compared to the novel context (though this difference only approached significance) [t(45) = 1.75] while there was a significantly faster RT for old nouns when presented in the novel context compared to the same-old context [t(45) = 2.75]. In Figure 16, it can be seen that, on "old" tests, the RT results of nonnouns are in the expected direction (same-old context faster than novel context) and the RT results of nouns are in the opposite direction than would be expected (novel context faster than the same-old context).

Insert Figure 16 about here

The analysis of the mean RTs for correct rejections revealed a significant effect of context condition [F(1,45) = 10.284, MSe = 31,323.2, p=.002]. The effect of word type [F(1,45)<1, MSe = 30,567.9] and the word type by context condition interaction [F(1,45)<1, MSe = 21,408.8] did not reach significance. These results are evidenced by the faster RT for "new" items presented in the novel context compared to the same-old context but no difference between nouns and nonnouns.

Experiments 1 through 5 concentrated on the same stimulus manipulation of HF versus LF words. Since the mirror effect can be found with a number of other stimulus manipulations, it was necessary to determine the generality of the findings discussed thus far. Accordingly, Experiment 6 compared the recognition of nouns and nonnouns in same-old and novel contexts. The pattern of hits and correct rejections in Experiment 6 were the same as was found in Experiments 3A, 3B, and 4 in that both the mirror effect and effects of environmental context were observed without evidence of any interaction. Again, this further supports the notion of separate processes involved in these two distinct effects.

A' and B''_D values were calculated for Experiment 6 as well. Analysis of A' values revealed an effect of word type, indicating participants' increased sensitivity to nouns as opposed to nonnouns. This finding is comparable to the LF advantage over HF found in Experiments 1 through 3. As this was the only significant effect of A' values, these findings are most similar to those of Experiment 3B that showed only the effect of word frequency.

Again showing similarities to Experiments 1 through 3, the B''_D values of Experiment 6 support the view that the decision criteria used by participants changes depending upon the context condition. There was a highly significant effect of context, indicating more conservative responses in the novel context condition in comparison to the same-old context condition. These findings, in addition to the first three experiments in the present study, provide strong evidence for the use of differing response criteria depending upon the context used at test.

When considering the results of the RT data of Experiment 6, findings across experiments become less consistent. For tests of novel items, an effect of context condition was found (novel context item recognized faster than same-old context), which is consistent with the RT results of the preceding experiments and the notion that context has a greater effect than word type/word frequency on tests of new items. Experiments 3A and 3B, on the other hand, also showed an effect of word frequency so it appears that word frequency produces more reliable effects on response latency for new items than does the difference between nouns and nonnouns. It is interesting to note that Experiment 6 was almost identical to Experiments 3A and 3B except that a different stimulus manipulation was used (i.e. nouns vs. nonnouns). The RT results for tests of old items failed to show an effect of word type, which was expected on the basis of the consistent effect of word frequency in the preceding experiments. While Experiment 3A demonstrated both an effect of word frequency and a frequency by context interaction, Experiment 6 revealed only a word type by context interaction and Experiment 3B showed only a main effect of frequency. The role of word type in the RTs of tests of old items in Experiment 6 can not be dismissed on account of the interaction that was found. However, in this case, word type (i.e. nouns vs. nonnouns) does not appear to have a greater effect upon RTs than context as was found with word frequency in the previous experiments.

Experiment 6 provided evidence that the majority of the results obtained using the stimulus manipulation of LF versus HF words in Experiments 1 through 5 generalize to another type of stimulus manipulation used to demonstrate the mirror effect (i.e. nouns versus nonnouns). The independence of the mirror effect and effects of environmental context for item recognition is well founded in the present study. However, the mirror effect has been found using the noun/nonnoun manipulation for word pairs (i.e. associative recognition) in addition to single words (Hockley, 1994). Experiment 7 addresses the question of whether the current findings will extend to tests of associative recognition memory by testing participants' recognition of word pairs instead of single words.

Experiment 7

Hockley (1994) not only found the mirror effect for recognition memory of item information using the noun-nonnoun manipulation, he also found the mirror effect for associative recognition (i.e. participants are tested on word pairs rather than single words). Due to the fact that participants obviously have had pre-experimental exposure to single items, item recognition not only requires the ability to discriminate "old" from "new" items during a recognition test, participants must also be able to temporally distinguish an item encountered at study from an item pre-experimentally encountered. Hockley (1992) pointed out that associative information presented to participants typically takes the form of a novel pairing of words that do not have a strong pre-experimental relation. Consequently, Hockley argued that recognition memory for item information may depend more upon context-related information than recognition memory for word pairs or associative memory. Murnane and Phelps (1993) were able to obtain an effect of context for associative recognition. However, in absolute terms, this effect seemed to be smaller than the effects they found for item recognition.

Thus far, the current study has demonstrated the statistical independence of the mirror effect and effects of environmental context. If associative recognition memory is in fact less reliant upon context-related information, it would be expected that extending the current experimental paradigm to recognition memory for word pairs would result in no effect of changing environmental context (or at least less of an effect) while still producing the mirror effect as was found by Hockley (1994). Experiment 7 will address this concern by replicating

Experiment 6 using old and new word pairs at test in place of single words. This way, it can be determined whether recognition memory for associations is less dependent on (or independent of) contextual information.

Method

Participants. Forty-four Wilfrid Laurier University undergraduate students were recruited from the Department of Psychology's participant pool. Students earned course credit for participation.

Stimuli and Apparatus. The stimuli and apparatus were identical to that used in Experiment 6.

Procedure. The procedure was identical to Experiment 6 except that participants were presented with 32 word pairs at study instead of the 24 used in Experiment 6. This was so the 16-word pair recognition test would contain 8 old word pairs that were the same pairings as during study and 8 novel word pairs whose words were presented at study but in a different pairing than presented at test (i.e. the individual words had been seen before but the association was novel) plus 4 primacy and 4 recency buffer pairs. There were 16 possible unique contexts, each of which were presented twice in the study list. Thus, at test participants made "new"/ "old" recognition decisions for 16 word pairs. Half of these word pairs were targets (from the study phase) and half were distractors. Of these "old" and "new" word pairs, half were nouns and half were nonnouns. Furthermore, half of the nouns and nonnouns were presented in the same-old context and the other half were presented in a novel context. The written instructions were modified so as to emphasize the importance of making associations between the words and to explain how to judge word pairs as "new" and "old".

Design. A 2 x 2 within-subjects factorial design was used for this experiment. The two factors were pair type (noun vs. nonnoun) and type of context (same-old vs. novel).

Results and Discussion

Analyses were conducted with mean hit rate, mean correct rejection rate, and A' and B''_D scores as dependent measures. Additionally, mean response time (RT) for hits and correct rejections were also analyzed. Means were calculated for each subject in each condition.

Figure 17 shows the proportion of correct responses for new and old tests for nouns and nonnouns in both context conditions. Noun pairs were recognized more accurately than nonnoun pairs, but this effect was smaller for new pairs than for old pairs. No consistent effects of environmental context were found for tests of old word or new word pairs. A 2 X 2 ANOVA based on hit rate revealed a highly significant effect of word type [$F(1,43) = 27.152$, $MSe = .019$, $p<.001$]. However, there was no effect of context condition [$F(1,43) = 3.780$, $MSe = .016$] and no word type by context condition interaction [$F(1,43) = 1.225$, $MSe = .017$]. The ANOVA for correct rejection rate did not show a significant effect of word type [$F(1,43) = 2.242$, $MSe = .017$] nor a significant effect of context condition [$F(1,43)<1$, $MSe = .011$]. Additionally, there was no significant word type by context condition interaction [$F(1,43) = 1.915$, $MSe = .015$] for correct rejections.

Insert Figure 17 about here

Mean A' values are presented in Table 8 for nouns and nonnouns in the novel and same-old context conditions. A 2 X 2 ANOVA of A' revealed a significant effect of word type [$F(1,43) = 10.500$, $MSe = .011$, $p=.002$] while the main effect of context condition [$F(1,43) =$

2.043, MSe = .006] and the word type by context condition interaction [F(1,43)<1, MSe = .005] did not approach significance.

Insert Table 8 about here

Mean B''_D values are presented in Table 8 for nouns and nonnouns in the novel and same-old context conditions. A 2 X 2 ANOVA of B''_D showed no significant effects of context condition [F(1,43)<1, MSe = .215], word type [F(1,43) =1.356, MSe = .197, ns] nor a word type by context condition interaction [F(1,43) =1.613, MSe = .209, ns].

The mean RTs for correct new and old responses in each condition are displayed in Figure 18. A 2 X 2 ANOVA was performed separately for hits and correct rejections. The analysis of the mean RTs for hits revealed a significant effect of word type [F(1,43) = 8.779, MSe = 282,329.4, p=.005] but no effect of context condition [F(1,43) = 1.914, MSe = 108,495.6, ns]. The word type by context condition interaction was also found to be significant [F(1,43) = 4.505, MSe = 228,489.0, p=.040]. Paired sample t-tests for these RT values showed that this interaction was a result of no difference in RT for old nonnouns presented in same-old and novel context conditions [t(45) = 0.86] while there was a significantly faster RT for old nouns when presented in the novel context compared to the same-old context [t(45) = 2.93]. In Figure 18, it can be seen that, on “old” tests, the RT results of nonnouns, while not significantly different, are in the expected direction (same-old context faster than novel context) and the RT results of nouns are in the opposite direction than would be expected (novel context faster than the same-old context).

Insert Figure 18 about here

The analysis of the mean RTs for correct rejections revealed a significant effect of context condition [$F(1,43) = 5.496$, MSe = 208,276.9, $p=.024$] and a significant effect of word type [$F(1,43) < 5.406$, MSe = 177,749.6, $p=.025$]. The word type by context condition interaction did not reach significance [$F(1,43)<1$, MSe = 171,417.8]. These results are evidenced by the faster RT for "new" items presented in the novel context compared to the same-old context and faster RTs for nouns compared to nonnouns.

In contrast to the previous experiments in the current study, Experiment 7 resulted in a drastically different pattern of findings in terms of hits and correct rejections. Among all proportion correct data, only a main effect of word type was found to be significant for hits (there was no effect of word type for correct rejections). The effect of word type for correct rejections was in the expected direction, but was not statistically reliable. More importantly, no effect of context condition was found at all for either hits or correct rejections. The results for hits support the suggestion that associative recognition memory relies less upon contextual information than does item recognition since the effect of word type was witnessed without an effect of context.

A' values in Experiment 7 again showed a main effect of word type, indicating participants' increased sensitivity to nouns as opposed to nonnouns. The stability of this effect over the present set of experiments demonstrates the underlying ability of participants to discriminate between different word types to produce effect of word type or word frequency (which leads to the mirror effect).

Throughout the present study, accompanying the main effects of context condition for the proportion correct data, a main effect of context has been found for B''_D values, suggesting the use of different decision criteria by participants when environmental context changes at test. However, in the present experiment, no effect of context condition was found for hits and correct rejections. Consequently, no effect of context condition was found in the analysis of B''_D. These results, considered together with the results of the previous experiments, provide strong support for the suggestion that effects of context are driven by differing response criteria used by participants when the context condition changes from study to test. If, as Hockley (1992) suggests, context plays less of a role in associative memory than item memory, then likely participants do not tend to use differing response criteria with associative recognition memory as context does not influence decisions to the extent as with item information.

Although substantially longer overall due to the nature of the experiment, RT results in Experiment 7 resembled those found in Experiment 3A. For tests of novel word pairs, an effect of context was found along with an effect of word type. Over the entire study, only effects of context were found consistently for "new" tests indicating a faster response time for new targets presented in a novel context. For tests of old word pairs, an effect of word type was consistent with Experiment 3A, 3B, and 5's effect of word frequency. However, Experiment 6 showed no effect of word type but demonstrated a word type by context condition interaction, which was also observed in Experiment 7.

General Discussion

The present study reports seven experiments that examined the effects of changing environmental context and the mirror effect in recognition memory. Using word frequency as an indicator of item strength, context was manipulated in such a way as to examine the effect of changing environmental context on the recognition of "new" and "old" items (Experiments 3A and 3B). In addition to this basic procedure, the importance of the appearance of a context as novel or same-old was assessed using a "rearranged-old" context in Experiment 1, and the degree of similarity of the learning and test context was manipulated in Experiment 2. Confidence judgments were obtained in Experiment 4 to determine the differences in confidence level across different contexts and word frequencies. Experiment 5 repeated some contexts during learning while presenting others only once to investigate how manipulating context strength would affect recognition decisions. A noun versus nonnoun strategy was also adopted to assess the generality of the observed results (Experiment 6) and to examine how changing environmental context would affect the mirror effect for associative recognition memory.

Over the seven experiments reported here, strong evidence of the statistical independence of effects of environmental context and the mirror effect (i.e. effects of word frequency and word type) in recognition memory were obtained. The pattern of hits and correct rejections in Experiments 1 through 6 demonstrated main effects of word frequency (word type in Experiment 6) and context condition but never showed an interaction between these two main effects. These results support the notion of two distinct processes that produce these individual effects. Further support for this idea is found in the analyses of A’ and B''_D values. The A’ results observed in the present study demonstrate participants’ consistent ability to discriminate between word frequencies/word types and is supported by a memory-based view of the mirror effect while the B''_D data consistently implies the presence of differing response criteria when context changes from study to test, which suggests the role of response bias in context effects. Due to the obvious distinction between the mirror effect and effects of environmental context for the proportion correct data, they will each be discussed separately with consideration of the views and findings of other researchers and the implications of the present experiments on these views.

<u>**The Mirror Effect (Effects of Word Frequency and Word Type)**</u>

The mirror effect for item recognition memory has been well demonstrated in the present study. Over the first five experiments, a significant effect of word frequency was found, showing the better performance on low-frequency words over high-frequency words characteristic of the mirror effect (this pattern of findings was also found for nouns over nonnouns in Experiment 6). While these findings have been demonstrated before (e.g. Glanzer & Adams, 1985; Hockley, 1994), the fact that the mirror effect can be observed to occur apparently independent of the effects of environmental context has important implications for the basis of this effect.

Hirshman and Arndt (1997) contrasted two competing theories of the mirror effect; a memory-based explanation and a decision-based explanation. According to the memory-based view of the mirror effect, different types of words have different levels of memory strength whether or not they are new or old and it is these differences in memory strength about a single decision criterion (assumed to be set between the old and new memory-strength distributions) that creates the mirror effect. This idea can easily be envisioned for "old" items (items originally seen in the learning phase). Low-frequency old items would have certain properties that increase their memory strength over high-frequency old items, making recognition of these items as "old" more likely. A similar process would occur for "new" items (items not seen in the study list). Low-frequency new items would have properties that lead to better discrimination from study list items than high-frequency new items would thus making LF new items more likely to be classified as "new" than high-frequency words.

However, the decision-based explanation suggests a different basis of the mirror effect. According to this hypothesis, while the low-frequency old items still have a greater overall strength than the high-frequency old items, low and high-frequency new items are represented together in the same memory strength distribution (neither is greater in memory strength). It is then suggested that the greater memory strength of low-frequency old items creates a higher response criterion than the lower memory strength of high-frequency items. So, whereas the memory-based approach is based on the assumption that recognition performance is a result of the differential memory strength of low-frequency and high-frequency items (both new and old) about a single decision criterion, the decision-based approach posits that recognition performance in the mirror effect is a result of different decision criteria that arise for low-frequency and high-frequency old items. Hirshman and Arndt (1997) point out that each theory suggests something different in terms of the independence of hit rates and false alarm rates (or correct rejection rates). The memory-based approach holds that the differences in hit and false alarm rates arise independently for old and new items, respectively, while the decision-based approach indicates that these differences are dependent upon one another. Over seven experiments, Hirshman and Arndt demonstrated that false alarm rates (and consequently, correct rejection rates) for different word frequencies could be found to be independent of the hit rate effects of these items. These findings are in direct support of the memory-based explanation of the mirror effect and pose problems for the decision-based view, which suggests that hit and false alarm rates should be dependent upon one another.

The results of the present study are consistent with the view that the mirror effect in recognition memory is a memory-based phenomenon as opposed to a result of differing response criteria. The effects of context, on the other hand, appear to be the result of differing response criteria (this will be discussed in the next section). This would satisfactorily explain the statistical independence of the two effects over the first six experiments. Further support comes from the analyses of A' and B''_D values. Over the seven experiments, a consistent effect of word frequency/word type was found for A' values, indicating participants' greater sensitivity to low-frequency over high-frequency words, and nouns over nonnouns. However, there was no consistent evidence to support the notion that response criteria differed for low- versus high-frequency words or nouns versus nonnouns (i.e. B''_D scores were not consistently different when the mirror effect was observed). These findings provide further support of the memory-based view of the mirror effect in recognition memory. Interestingly, Experiment 7 of the present study found a main effect of word type with the analysis of hit rate but not with the analysis of correct rejections. Experiment 7 used word pairs instead of single words to test the effect of changing context on the mirror effect in associative recognition memory. Hirshman and Arndt (1997) demonstrated how the independence of correct rejection rate from hit rate supports a memory-based view of the mirror effect. In the present study, an effect of word type was found for hits whether the test probes were single items or word pairs. However, a significant effect of word type for correct rejections was found using single items but not word pairs. According to a decision-based model, a criterion shift that results in higher hit rates should also reduce false alarms (thus increasing correct rejections). As this is not what was observed to occur when word pairs were used in Experiment 7, the suggestion that the underlying memory distributions are independent is supported and the memory-based model of the mirror effect gains further endorsement.

Effects of Environmental Context

Three different explanations have been offered for the effects of changing environmental context in recognition memory tasks; the outshining hypothesis, global activation models, and the suggestion of response bias. The outshining hypothesis posits that, in a recognition test, both the item and the context it is presented in each have some degree of cue strength in memory (Murnane & Phelps, 1995). When the strength of the item is greater than that of the context, its memory trace "outshines" the strength of the context. These two memory cues are then suggested to compete when recognition decisions are made. According to the outshining hypothesis, the results of the present study should have been quite different. It would have been expected that changing the environmental context of a to-be-remembered (TBR) word would hinder performance on the weaker high-frequency words more than performance on the stronger low-frequency words. Another way to look at this prediction would be to say the mirror effect would be more pronounced when context is changed from study to test. This prediction was not supported by the present study. Over the seven experiments, no evidence was found to suggest that changing context differentially affects weaker items more than stronger items. While it has previously been shown by Smith et al. (1978) that recognition performance for LF words was not affected by changing context while performance on HF words was diminished, the experiment they reported used a between-subjects design that has been shown to be ineffective at demonstrating effects of environmental context. The present study's consistent lack of any interaction between context condition and item strength (i.e. word frequency) contradicts the predictions offered by the outshining hypothesis.

The prediction of changing context affecting HF words more than LF words (or nonnouns more than nouns) is not in agreement with the current global activation models discussed by Murnane and Phelps (1995). These models predict quite the opposite outcome. Global activation models suggest that an item and its context are integrated into the same memory cue and activate memory together in a more global manner. With this view, when an item is presented at test in the same context as it was presented in at study, activation would be greater than if it was presented in a different context (as the same context provides added memory activation). Murnane and Phelps found that as item strength was increased, the magnitude of the context effect either increased or remained the same. Taking word frequency/word type as a measure of item strength, it would be expected that LF words would be affected more than HF words by a change in environmental context. The present study seems to provide some difficulty for the global activation models of recognition memory. When Murnane and Phelps manipulated item strength, they did it in a way that would enhance the association between the item and its environmental context. For example, in one experiment they increased the number of presentations of some items in a specific context and in another they increased the length of time that some words were presented in their context for. It may be more correct to say that Murnane and Phelps were manipulating item-context associations. In this manner, it seems clear how increasing the strength of the association between item and context could lead to more of an effect of changing context from study to test.

The present study used a different measure of item strength with word frequency/word type manipulations in that the strength of an item was independent of its association to any particular context. The independence of item strength and context was apparent in the results of Experiments 1 to 6, which showed no differential effects between LF and HF words (also nouns and nonnouns in Experiment 6) when context was changed from study to test. Additionally, Experiments 2 and 5 sought to manipulate the strength of the contexts independent of the individual items. In addition to the usual same-old and novel contexts at test, Experiment 2 used intermediate contexts that consisted partially of old context elements and partially of new context elements. Global activation models would predict the highest amount of activation for items presented in the same-old context and the lowest amount of activation for items presented in a novel context with those presented in the intermediate contexts having activation correspondingly in the middle. While this pattern was found for LF words in Experiment 2, it was not apparent for HF words. Due to the ambiguous nature of these results, Experiment 5 manipulated context strength by repeating some contexts multiple times during study in a manner similar to Murnane and Phelps (1995) while presenting others only once (note that the study pairs were not repeated as in Murnane and Phelps' experiment). According to global activation models, the repeated context should provide greater activation when presented at test resulting in a higher hit rate for old words presented in a repeated context (compared to a single presentation context) and a lower correct rejection rate. However, absolutely no difference was found between items presented in a repeated context and those presented in a single presentation context. The ability of changing environmental context to differentially affect items of greater and lesser strength as predicted by global activation models seems only to be limited to instances when the strength of the item-context pairing is manipulated (as was accomplished in Murnane & Phelps, 1995).

Further problems for global activation models are found when findings in support of a third hypothesis are considered. Feenan and Snodgrass (1990) have suggested that changing environmental context from study to test does not have an effect on memory representations per se but rather change the participants' likelihood of identifying an item as "old" or "new". In a review of several studies (including their own), Feenan and Snodgrass demonstrated that a change in context from study to test resulted in participants adopting a more conservative response criterion (i.e. made them less likely to identify an item as "old"). A change in context affected participants' discrimination between same-old and novel context items in only one of their experiments. Feenan and Snodgrass concluded that, while changing context from study to test may have a small effect on discrimination, the major effect of changing context is a change in response bias. In simpler terms, an item that is presented in a same-old context appears more familiar to participants than an item presented in a novel context.

Murnane and Phelps (1994) have argued that subject-initiated changes in response criteria were not a factor in their studies for three reasons. Their first point was that test context was manipulated within subjects and within lists (with same and different contexts randomly intermixed at test) making it necessary for criterion shifts to occur on a trial-by-trial basis. Secondly, Murnane and Phelps point out that test context is not predictive of correct response. Lastly, they note that, when performance feedback is not given, participants would not have any external motivation to adjust their response criterion. In response to the first point, that criterion changes would have to occur on a trial-by-trial basis, since the difference in the same-old and novel context conditions is quite obvious, changing criterion from trial to trial does not seem unreasonable. Additionally, most mean response times were over one second in duration, which is more than enough time for criterion changes to occur. Secondly, while test context is not predictive of correct response, participants may not be aware of this, and even if they are they may use or be biased by contextual information anyway. Finally, although participants are not given any external motivation to adjust their response criterion, this does not mean that they are not using internal motivation. In many situations, participants have to set their own response criterion without feedback or external motivation.

The results of the present study parallel those of Feenan and Snodgrass (1990) exceptionally well. Primarily, all tests of item recognition in which A' and B''_D values were obtained (Experiments 1,2,3, and 6) revealed that participants adopted a more conservative response criterion for items that were presented in a novel context. Furthermore, while word frequency and word type manipulations had an effect on the discriminability of test items, environmental context failed to show any consistent effect on the discriminability of items.

One other source of support for Feenan and Snodgrass' response bias view is the results of Experiment 7 of the present study. It has been suggested that associative recognition memory relies less upon contextual information than does item information (Hockley, 1992). When participants were tested on word pairs instead of single words, hit and correct rejection rates were not affected by changes in context. Additionally, participants failed to show evidence of the conservative response bias that occurred when context was changed for recognition tests of item information (discriminability was also not affected by context). Therefore, Experiment 7 demonstrated a lack of criterion changes when environmental context had no effect on performance. These findings suggest that associative recognition is less susceptible to the biasing effects of environmental context than item recognition.

One other aspect of the current series of experiments that deserves to be mentioned here is the conditions under which the novel context condition can be effective in altering recognition performance of old items. Primarily, the results of Experiment 1 should be considered. In Experiment 1, two types of "different" contexts were used. One was a completely novel context made up of screen elements that had not been seen during the study phase. The other was a context composed of old context elements that were rearranged into a novel combination that had not been used during study. It is interesting to note that participants treated this "rearranged" context in an identical manner to the same-old context condition. These findings tend to suggest the importance of the "look" of a test context. For instance, the novel context holds much less similarity in appearance to the study contexts than the rearranged contexts in that no part of it was seen during study so it stands out as something completely new. The rearranged context, on the other hand, has the look of an old context. The lack of any discriminability between contexts on the part of participants would lead one to suggest that as long as the test context has the look of an old context, it will "feel" more familiar to participants, rather than there being any context-specific memory for items. This suggestion is supported further by the results of Experiment 5 that showed no difference in recognition performance between an item presented in a context seen eight times at study and one presented in a context presented only once during study. This further downplays the notion of context-specific memory.

While there have been several theories proposed to account for the effects of environmental context in recognition memory, the most salient view at this point is that offered by Feenan and Snodgrass (1990). The outshining hypothesis has been shown to be ineffective in explaining the effects of environmental context previously (Murnane & Phelps, 1995) and has not been able to account for the pattern of results observed in the present study. While the pattern of hits and correct rejections in the present set of experiments could be argued to be explained by global activation models, this theory loses strength when one considers the changes in the estimated response criteria associated with changes in context. This, in addition to the lack of effect changing context has on the discriminability of test items, provides remarkable support for the belief that context effects are primarily a result of a response bias (i.e. a more conservative response criterion) that occurs when context changes from study to test.

Implications of the Response Time Data

The present study uncovered a very consistent pattern of results for the proportion correct data (i.e. hit and correct rejection rates). However, another pattern was found for the response time data, albeit less consistent than that of the proportion correct data. For the experiments in which response time was measured (i.e. every experiment except for Experiment 4), a different pattern arose for hits and correct rejections. For hits, word frequency/word type had a more consistent effect than did environmental context (Experiment 6 was the only experiment in which an effect of the stimulus manipulation did not occur although an interaction with context was observed). However, correct rejections seemed to be more susceptible to the effects of environmental context as this effect was consistently found whereas an effect of word frequency/word type was found in only about half of the experiments.

Essentially, the experiments of interest when considering response time data are only those after Experiment 2. It became apparent from Experiment 2 that the time it took participants to locate words presented in "old" locations (i.e. in one of the corners of the computer screen) was influencing the pattern of response times. However, it is interesting to note that after location was removed as a context element and all words were henceforth presented in the centre of the computer screen, the consistent patterns for hits and correct rejections for the remainder of the experiments matched the effects found in Experiment 1 (when screen location was still a context element). That is, an effect of word frequency was found for hits and not correct rejections while an effect of context was found for correct rejections and not hits. The reason that word frequency/word type has a greater effect on hits and environmental context has a greater effect on correct rejections is unclear. Hockley (1994) found in an investigation of the mirror effect that response times also produced the mirror effect in addition to proportion correct data (i.e. old and new low-frequency words were identified faster than old and new high-frequency words). An obvious question is why this effect of frequency becomes less reliable when context is manipulated. One could speculate from the preceding discussions of the mirror effect and effects of environmental context that the consistent effect of context for correct rejections is influenced primarily by decision-based processes while the consistent effect of word frequency/word type for hits is reflective more of memory-based processes. While it is obvious that both stimulus and context manipulations play a role in determining both hit and correct rejection response times, the reason that each has a greater effect under different conditions is yet to be determined. It could be suggested that effects of word-frequency, a result of memory-based recognition processes, has a dominant effect on the response time for tests of old items, when the test item has been seen before. Response times for tests of new items, on the other hand, may be more affected by changes in context when the test items had not been seen during study, thereby being influenced more by decision-based recognition processes. However, this is only speculation and the specific mechanisms and conditions whereby this pattern would result are impossible to determine without a full, predictive model of response times for manipulations of word frequency/word type and manipulations of environmental context. At present, such a model does not exist.

Suggestions from Studies of the Revelation Effect

Another area of study in recognition memory that has been receiving considerable interest in recent years is the revelation effect. The revelation effect refers to the finding that participants will be more likely to identify an item as being "old" (from a study list) if that item is presented to them in some distorted manner and somehow revealed (i.e. by the solving of an anagram) just before a recognition decision is required. This will occur both for targets (i.e. words from the study list) and for distractors (i.e. words not from the study list). Of interest to the present discussion is the fact that many recent findings concerning the revelation effect are comparable to the findings concerning the effects of environmental context in the present set of experiments.

In a recent study on the revelation effect by Cameron and Hockley (under review), two interesting results that parallel certain aspects of the effects of environmental context in the present study were found. First, it was found that the revelation effect can occur even when the word that is revealed to a participant is different from the word they must provide a recognition decision for, and the magnitude of this effect is similar to when the revealed word and the test word are the same showing that the revelation effect is not item-specific. In Experiment 1 of the present study, it was found that a rearranged context (novel combination of old context elements) provided the same advantage for old items over a novel context as did the same-old context, showing the effects of context are not context-specific. Thus, it appears that revealing an item prior to a recognition decision and presenting an item in an old-looking context provide an advantage for old items (and disadvantage for new items) that is not a result of memory-based processes rather, these procedures tend to make items seem more familiar to participants.

Additionally, Cameron and Hockley (under review) found that, given ample encoding time, the revelation effect did not occur for associative recognition memory. Similarly, in Experiment 7 of the present study, effects of environmental context were not found when tests of associative recognition memory were performed. Cameron and Hockley suggested that these findings resulted from associative recognition being based on a recall-like process. To test this, they substantially reduced presentation time in their study list to reduce the opportunity for recall and found that associative recognition was susceptible to the revelation effect under these conditions. While it is interesting to note that associative recognition is less susceptible to both the revelation effect and effects of environmental context, it would be worthwhile to reduce the study time in a similar manner for associative recognition in this study's experimental procedure to see if effects of context would occur under conditions where recall is less likely to occur.

Most significant to the present study is the fact that it has been shown that the magnitude of the revelation effect does not differ between low-frequency and high-frequency words (Hicks & Marsh, in press). The two most prominent explanations of the revelation effect are similar to those of the effects of environmental context in that one focuses on global activation models of memory (Westerman and Greene, 1998) and the other, offered by Hicks and Marsh, involves changing response criteria so that after an item is revealed, participants adopt more liberal response criteria (analogous to a same-old context condition), and more conservative response criteria are used when an item is not revealed beforehand (analogous to a novel context condition). Considering the arguments discussed by Hirshman and Arndt (1997), the independence of the revelation effect from the effects of word frequency (i.e. the mirror effect) may result from processes similar to the effects of environmental context. In fact, the view proposed by Hicks and Marsh suggest that the revelation effect is due to differences in response criteria across the different conditions, much in the same manner that Feenan and Snodgrass (1990) suggest occurs for the effects of environmental context. Further study is required to solidify these arguments but it would seem that the answers to whether the effects of context are a memory-based or a decision-based phenomenon may have important implications for the study of the revelation effect, and vice versa.

Concluding Statements

The present study has added to the growing amount of research on the mirror effect and the effects of environmental context in recognition memory by investigating both effects within one experimental design. By testing high- against low-frequency words and nouns against nonnouns in varying same-old and different context conditions, the statistical independence of the mirror effect and effects of environmental context was thoroughly established for hit and correct rejection rates. These results were shown not to be an artifact of differing confidence levels across the different conditions as the same results were found conditionalized on high-confident responses (see Experiment 4). This independence of effects leads to the suggestion of different processes underlying each of these processes in recognition memory. The evidence gathered from these experiments supports the notion suggested by Feenan and Snodgrass (1990), which holds that effects of environmental context arise from a response bias generated from the differing placement of response criteria across different context conditions. While criterion change is a major factor in the effects of context, it seems not to play a role in the mirror effects observed in the present experiments. Conversely, the consistent ability of participants to discriminate between word frequencies and word types supports a memory-based explanation that posits that different word types have characteristics that make them more or less discriminable whether they are old (from a study list) or new (distractors not from a study list). Similarities of this study's findings concerning effects of context and the reports of other studies concerning the revelation effect (e.g. Cameron & Hockley, in press; Hicks & Marsh, in press) imply that the independence of the memory-based mirror effect and criterion change-based context effects is not an isolated occurrence.

The similarities between context effects and the revelation effect should be further examined to clarify the roles of changing criteria in recognition memory. While the similarities noted here are interesting they are so only in that similar processes may be at work but to what extent they are similar cannot be determined at this point. Employing a revelation effect procedure within same-old and novel context conditions may serve to augment these effects, or perhaps one may dominate over the other. While these manipulations appear to shift response criteria, their effectiveness in the presence of other criterion-changing phenomena is unclear at present. Additionally, further study is necessary to elaborate on the inability of context to have an effect on associative recognition memory. Murnane and Phelps (1993) found an effect of context using a similar procedure as the one employed in the present experiment. However, the magnitude of the context effects for associative information was much less than that of item information. While both studies support the notion that context has less of an effect on associative information, a procedure that tests both item and associative information simultaneously would serve to elucidate this discrepancy.

References

Cameron, T. E., & Hockley, W. E. (1998). *The revelation effect for item and associative recognition: Familiarity versus recollection.* Manuscript under review.

Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society, 31,* 271-274.

Feenan, K., & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition, 18,* 515-527.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation, 14,* 375-399.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 5-16.

Glanzer, M., Adams, J. K., & Iverson, G. (1991). Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 81-93.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66,* 325-331.

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology, 71,* 99-104.

Hicks, J. L., & Marsh, R. L. (in press). The revelation effect: Anything more than a decrement to familiarity resulting in a liberal criterion shift in recognition memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 275-289.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1306-1323.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321-1330.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22,* 713-722.

Kim, K., & Glanzer, M. (1994). Attention/Likelihood theory: Reply to Hintzman (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 206-208.

Kucera, F., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence RI: Brown University Press.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 882-894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition, 22,* 584-590.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 158-172.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thompson (Eds.), *Memory in context: Context in Memory* (pp.13-34). New York: Wiley.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342-353.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Footnote

[1] A' and d' are both estimates of recognition performance that are hypothetically independent of decision criteria. A' is an indicator of discriminability in that it measures a participant's ability to remember and discriminate between old and new items on a recognition test. A' values can vary from 0 to 1 (0.5 representing chance performance) and is equivalent to proportion correct on a 2-alternative forced-choice recognition test. B''_D is the measure of the response criterion associated with A' (Donaldson, 1992). It refers to the placement of a decision criterion along a distribution which influences a participant's willingness to identify a test item as "old" (whether that item is in fact from the study list). B''_D values can range from -1 to +1, positive values representing conservative responding while negative values indicate liberal responding.

Table 1

Mean d', A', and B"D Values for High and Low Frequency Words in the New, Same Old, and Rearranged Old Context Conditions in Experiment 1.

	Low Frequency			High Frequency		
	Same/Old	Rearranged	Novel	Same/Old	Rearranged	Novel
d'	2.19	2.22	2.45	1.67	1.73	1.77
A'	.897	.882	.895	.837	.840	.828
B"D	.006	-.008	.449	.165	.090	.452

Table 2

Mean A' Values for High and Low Frequency Words in the Same Old, Old Colour/New Location, New Colour/Old Location, and New Context Conditions in Experiment 2.

	Same/Old	Old Colour	Old Location	Novel	**Mean**
HF	.837	.836	.840	.834	**.837**
LF	.884	.869	.869	.882	**.876**
Mean	**.860**	**.852**	**.855**	**.858**	

Table 3

Mean B"D Values for High and Low Frequency Words in the Same Old, Old Colour/New Location, New Colour/Old Location, and New Context Conditions in Experiment 2.

	Same/Old	Old Colour	Old Location	Novel	**Mean**
HF	.310	.174	.292	.406	**.296**
LF	-.001	.223	.211	.407	**.210**
Mean	**.155**	**.199**	**.251**	**.406**	

Table 4

Mean A' Values for High and Low Frequency Words in the Same Old and Novel Context Conditions in Experiment 3A and 3B.

	Experiment 3A			Experiment 3B		
	Same-old	Novel	**Mean**	Same-old	Novel	**Mean**
HF	.860	.880	**.870**	.810	.840	**.825**
LF	.900	.910	**.905**	.880	.890	**.885**
Mean	**.880**	**.895**		**.845**	**.865**	

Table 5

Mean B''_D Values for High and Low Frequency Words in the Same Old and Novel Context Conditions in Experiment 3A and 3B.

	Experiment 3A			**Experiment 3B**		
	Same-old	Novel	**Mean**	Same-old	Novel	**Mean**
HF	-.110	.380	**.135**	.070	.310	**.190**
LF	-.050	.260	**.105**	.070	.410	**.240**
Mean	**-.080**	**.320**		**.070**	**.360**	

Table 6

Mean Confidence Judgments for High and Low Frequency Words in the Same Old and Novel Context Conditions on Tests of Old and New Items in Experiment 4.

	NEW TESTS			OLD TESTS		
	LF	HF	**Mean**	HF	LF	**Mean**
Same-Old	2.417	2.267	**2.342**	2.613	2.739	**2.676**
Novel	2.474	2.303	**2.419**	2.554	2.682	**2.618**
Mean	**2.446**	**2.285**		**2.584**	**2.711**	

Table 7

Mean A' and B"D Values for Nouns and Nonnouns in the Same Old and Novel Context Conditions in Experiment 6.

	A' Values			B"D Values		
	Same-old	Novel	**Mean**	Same-old	Novel	**Mean**
Nonnouns	.841	.839	**.840**	.027	.326	**.177**
Nouns	.869	.870	**.870**	.033	.403	**.235**
Mean	**.855**	**.855**		**.030**	**.365**	

Table 8

Mean A' and B"D Values for Noun and Nonnoun Word Pairs in the Same Old and Novel Context Conditions in Experiment 7.

	A' Values			**B"D Values**		
	Same-old	Novel	**Mean**	Same-old	Novel	**Mean**
Nonnouns	.801	.791	**.796**	.166	.268	**.217**
Nouns	.858	.836	**.847**	.175	.103	**.139**
Mean	**.830**	**.814**		**.171**	**.186**	

Figure Captions

Figure 1. Mean proportion of correct responses for low (LF) and high frequency (HF) new and old tests in the same-old, rearranged-old, and novel context conditions of Experiment 1.

Figure 2. Mean response time (RT) in milliseconds of responses for LF and HF new and old tests in the same-old, rearranged-old, and novel context conditions of Experiment 1.

Figure 3. Mean proportion of correct responses for LF and HF new and old tests in the same-old, old colour, old location, and novel context conditions of Experiment 2.

Figure 4. Mean RT in milliseconds of responses for LF and HF new and old tests in the same-old, old colour, old location, and novel context conditions of Experiment 2.

Figure 5. Mean proportion of correct responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 3A.

Figure 6. Mean RT in milliseconds of responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 3A.

Figure 7. Mean proportion of correct responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 3B.

Figure 8. Mean RT in milliseconds of responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 3B.

Figure 9. Mean proportion of correct responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 4.

Figure 10. Mean proportion of high-confidence correct responses for LF and HF new and old tests in the same-old and novel context conditions of Experiment 4. Note the different scale used due to the smaller proportions being displayed.

<u>Figure 11.</u> Mean proportion of correct responses for LF and HF 1x and 8x study context condition old tests in the same-old and novel context conditions of Experiment 5.

<u>Figure 12.</u> Mean proportion of correct responses for LF and HF new tests in the 1x old, 8x old, and novel context conditions of Experiment 5.

<u>Figure 13.</u> Mean RT in milliseconds of correct responses for LF and HF 1x and 8x study context condition old tests in the same-old and novel context conditions of Experiment 5.

<u>Figure 14.</u> Mean RT in milliseconds of correct responses for LF and HF new tests in the 1x old, 8x old, and novel context conditions of Experiment 5.

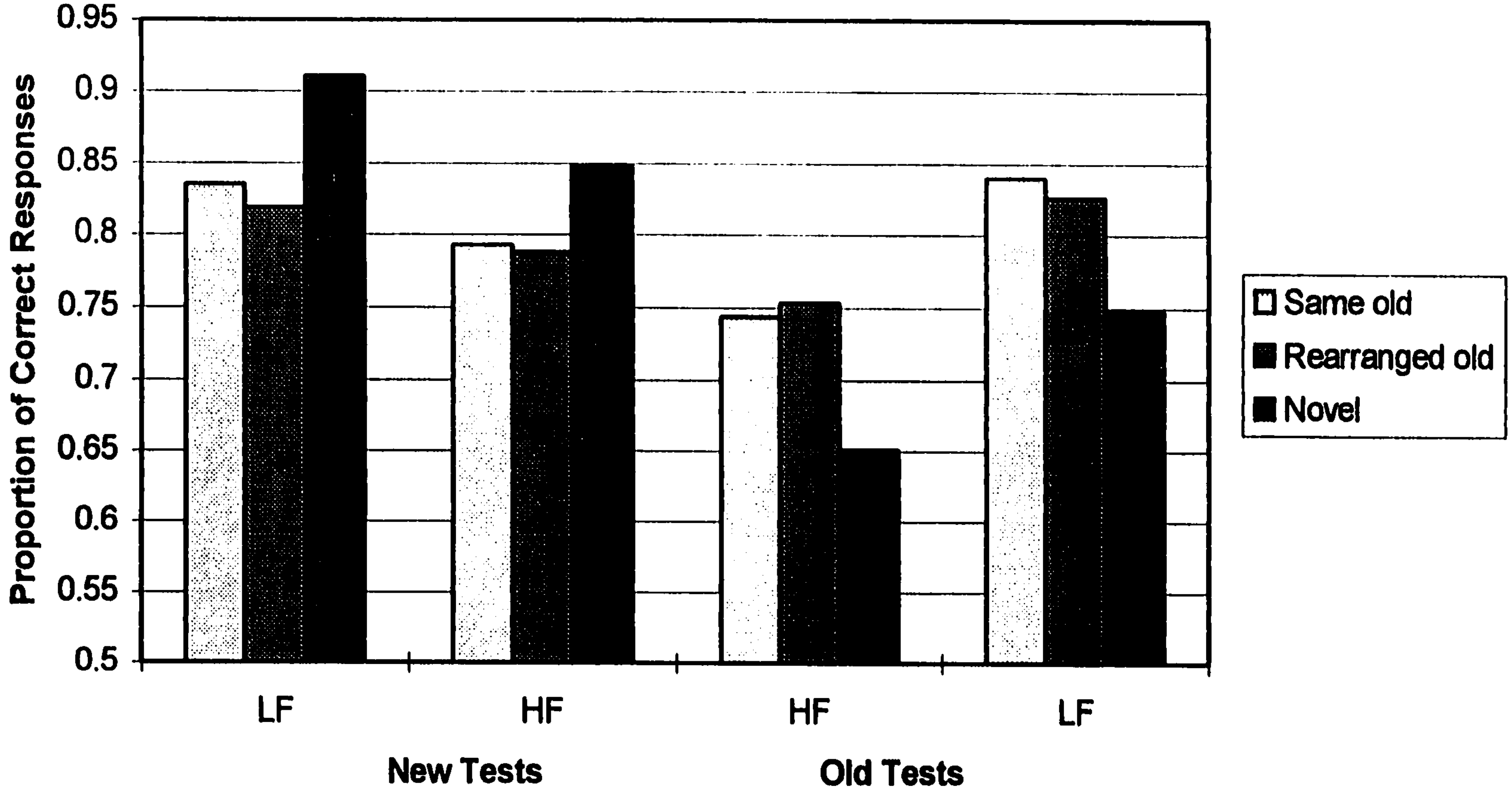
<u>Figure 15.</u> Mean proportion of correct responses for noun and nonnoun new and old tests in the same-old and novel context conditions of Experiment 6.

<u>Figure 16.</u> Mean RT in milliseconds of responses for noun and nonnoun new and old tests in the same-old and novel context conditions of Experiment 6.

<u>Figure 17.</u> Mean proportion of correct responses for noun and nonnoun word pair new and old tests in the same-old and novel context conditions of Experiment 7.

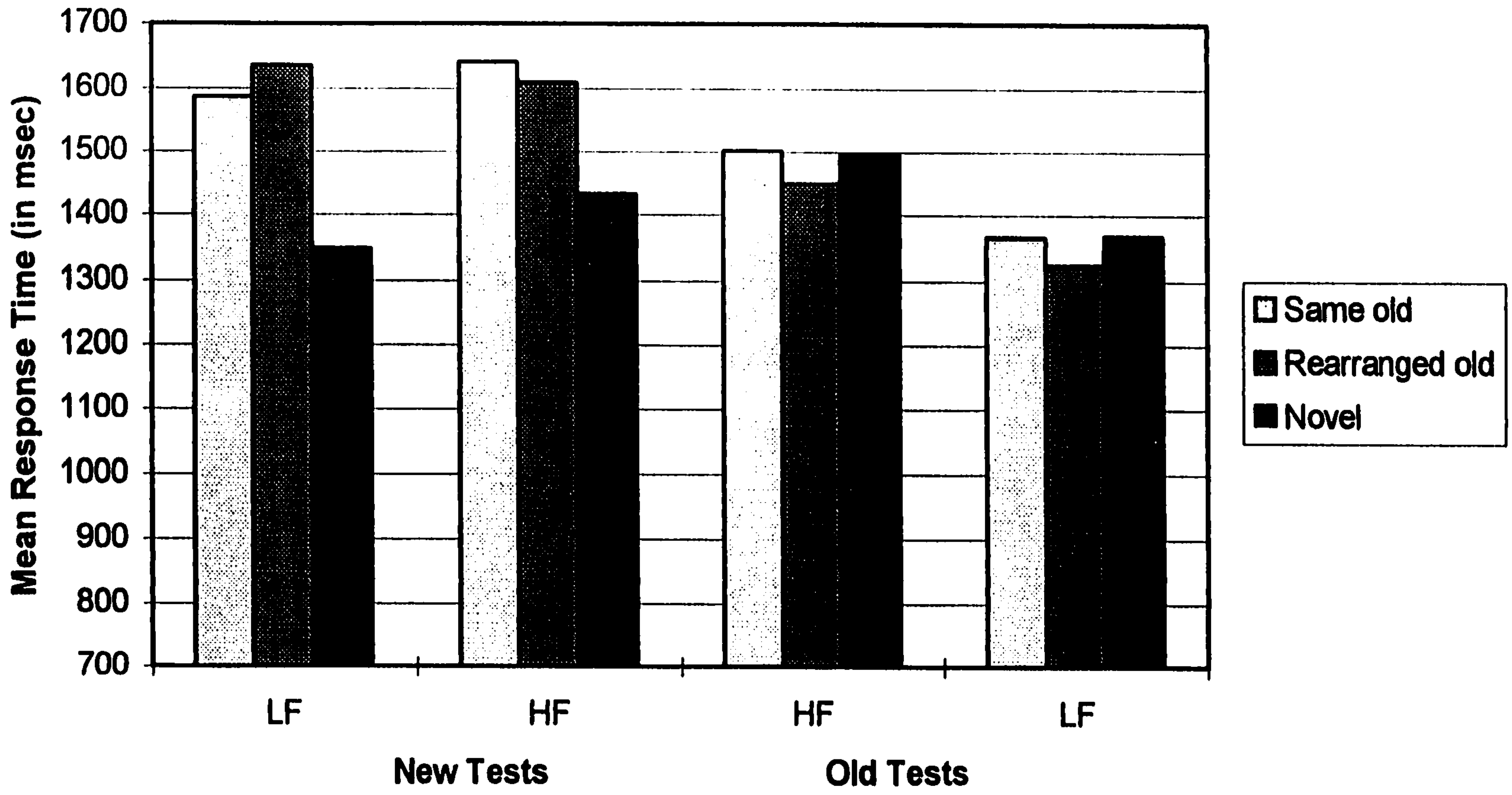
<u>Figure 18.</u> Mean RT in milliseconds of responses for noun and nonnoun word pair new and old tests in the same-old and novel context conditions of Experiment 7. Note the different scale used due to longer mean RTs for associative recognition.

Figure 1



Proportion of Correct Responses
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
LF
HF
HF
LF
New Tests
Old Tests
Same old
Rearranged old
Novel

Figure 2



1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
Mean Response Time (in msec)
LF
HF
HF
LF
New Tests
Old Tests
Same old
Rearranged old
Novel

Figure 3

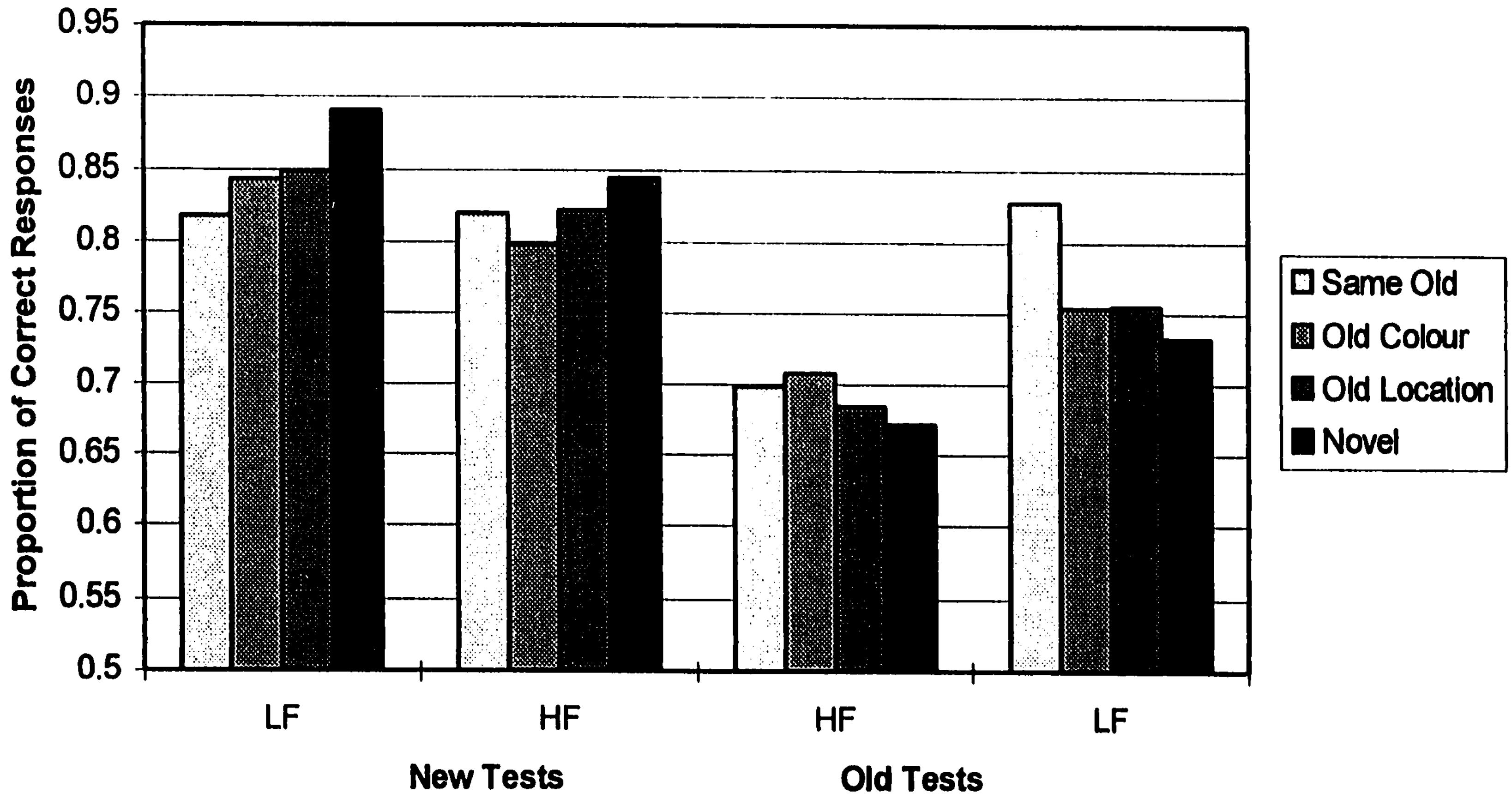

0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Proportion of Correct Responses
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Old Colour
Old Location
Novel

Figure 4

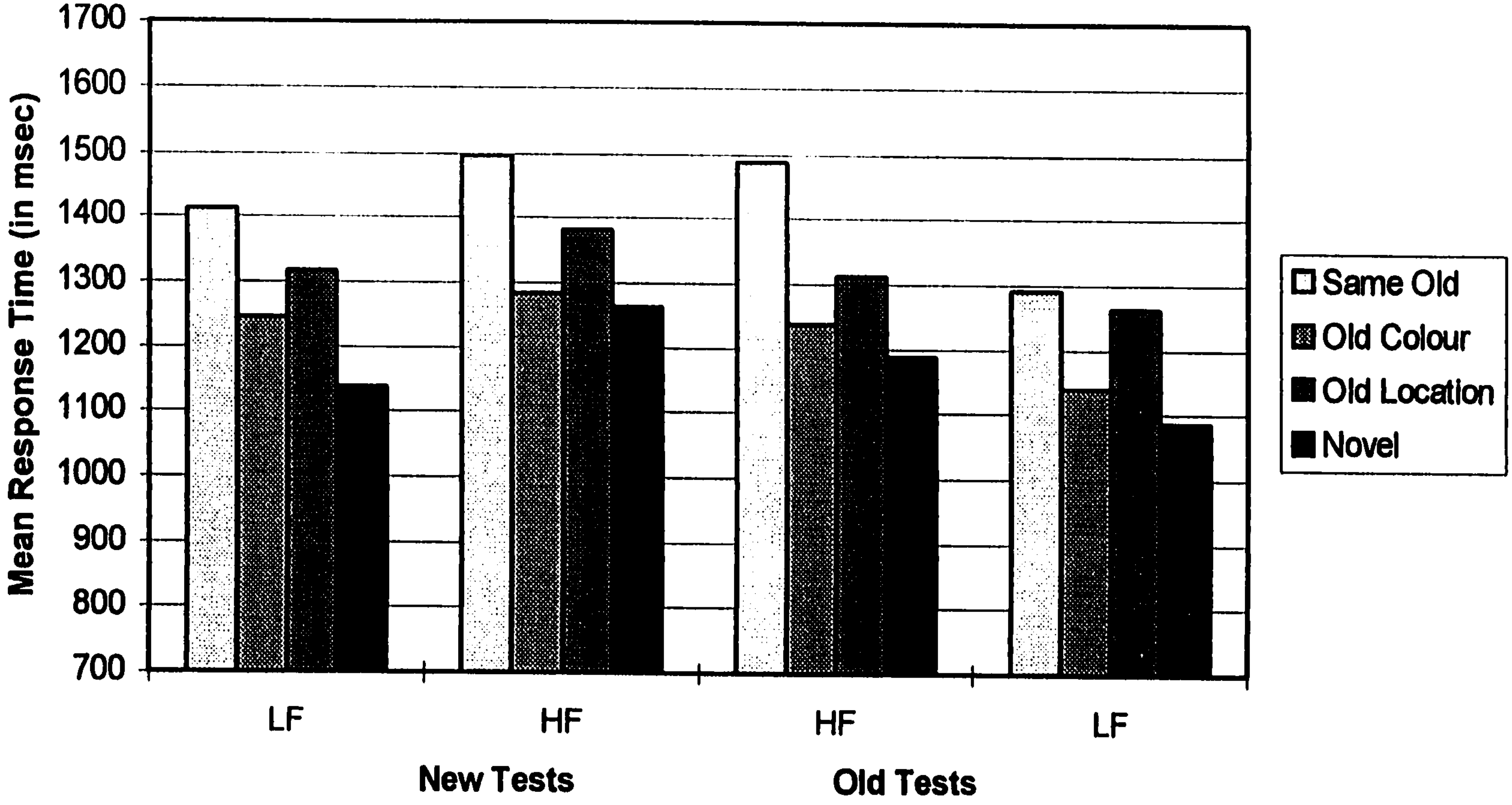

Mean Response Time (in msec)
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Old Colour
Old Location
Novel

Figure 5

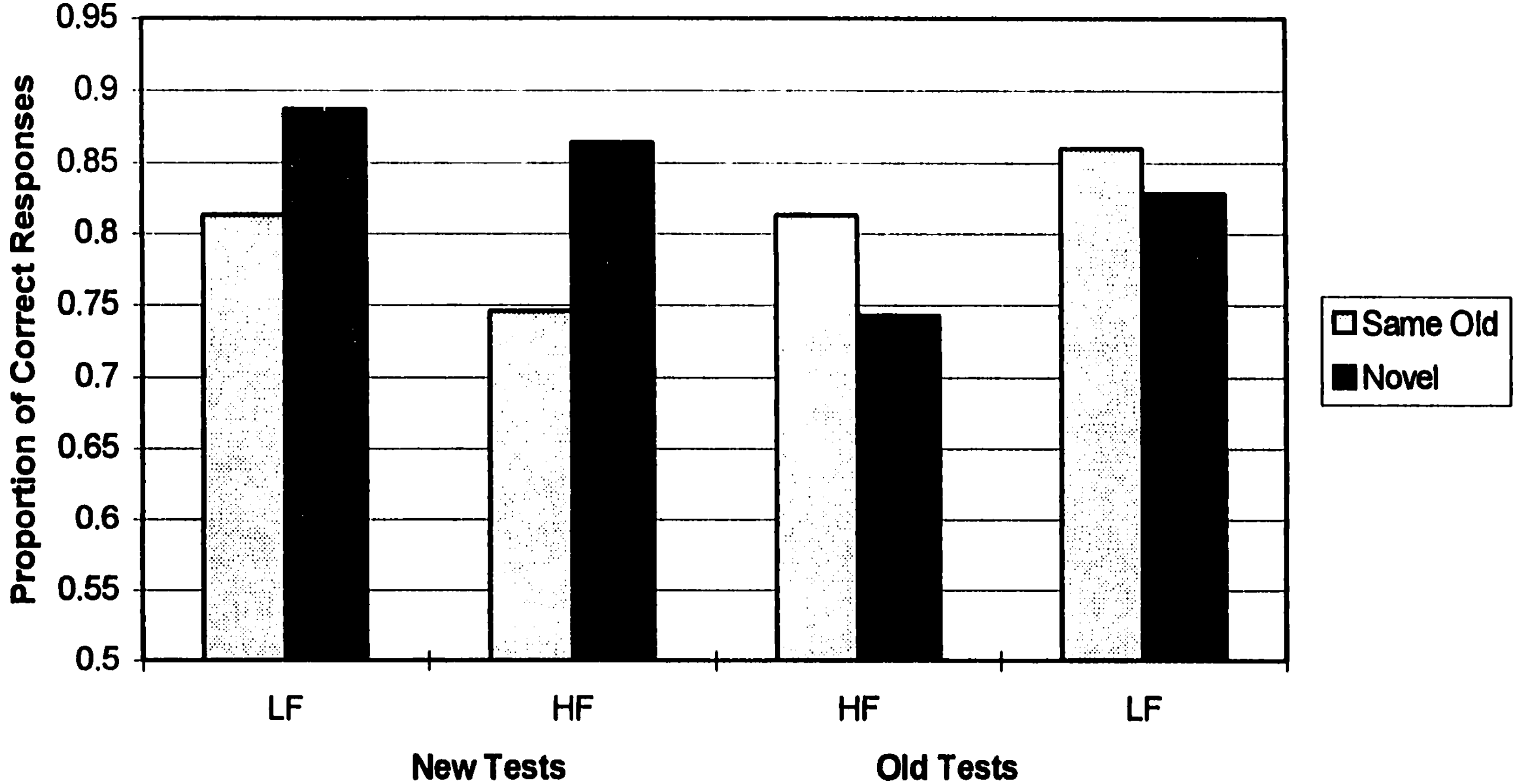
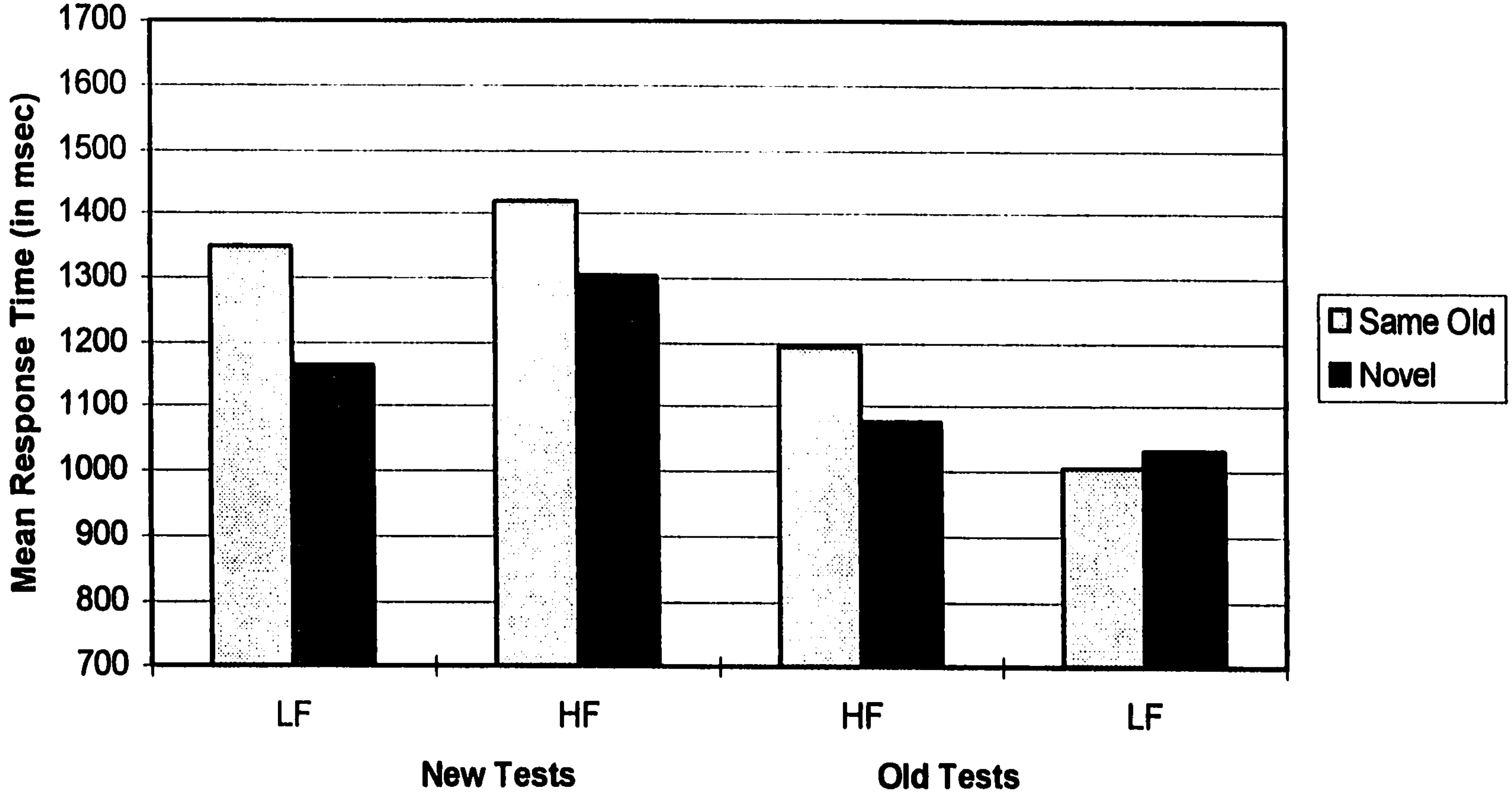

0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Proportion of Correct Responses
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Novel

Figure 6

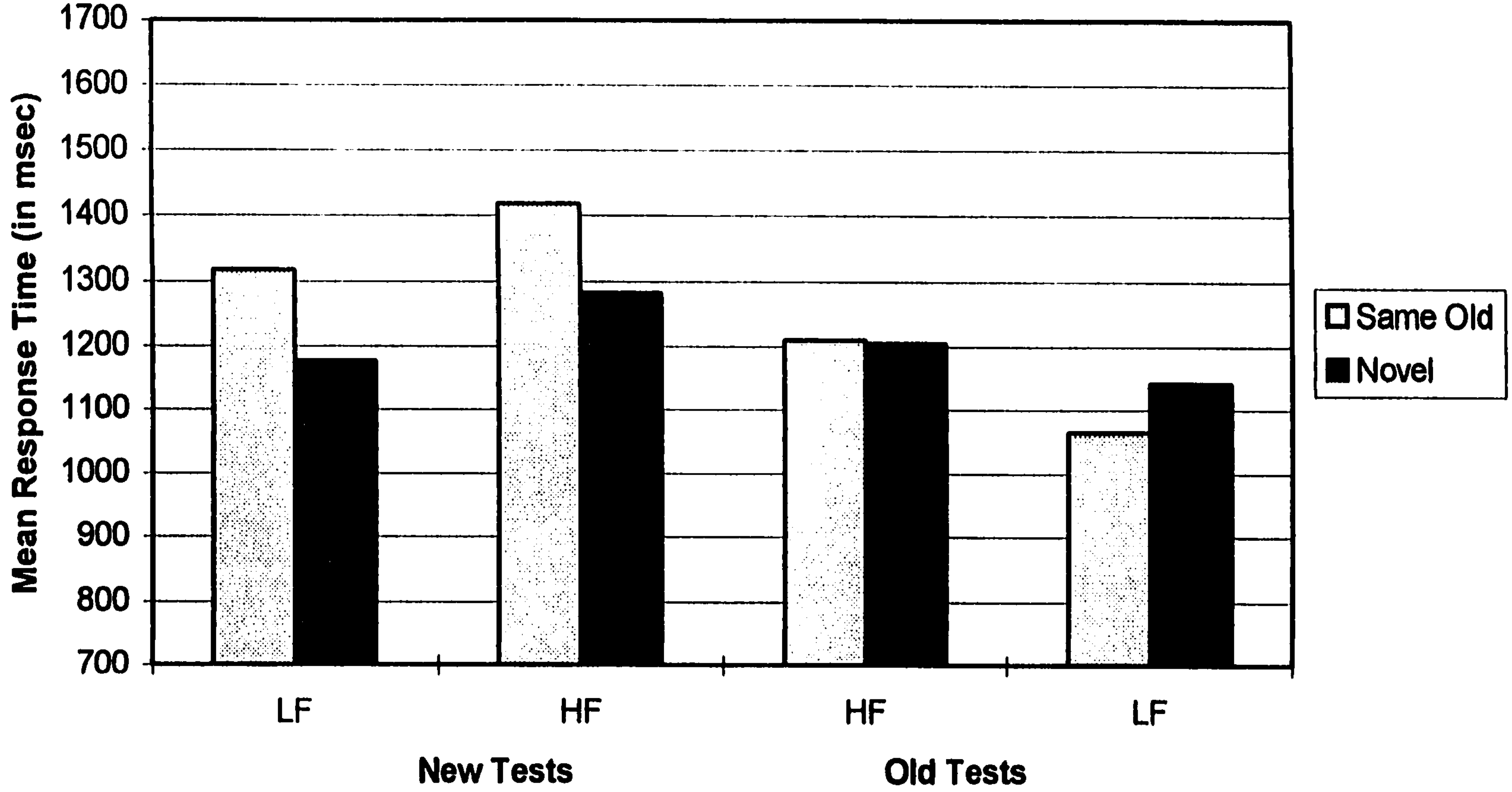


Mean Response Time (in msec)
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Novel

Figure 7


Same Old
Novel
Proportion of Correct Response
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
LF
HF
HF
LF
New Tests
Old Tests

Figure 8



Mean Response Time (in msec)
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
Same Old
Novel
LF
HF
HF
LF
New Tests
Old Tests

Figure 9

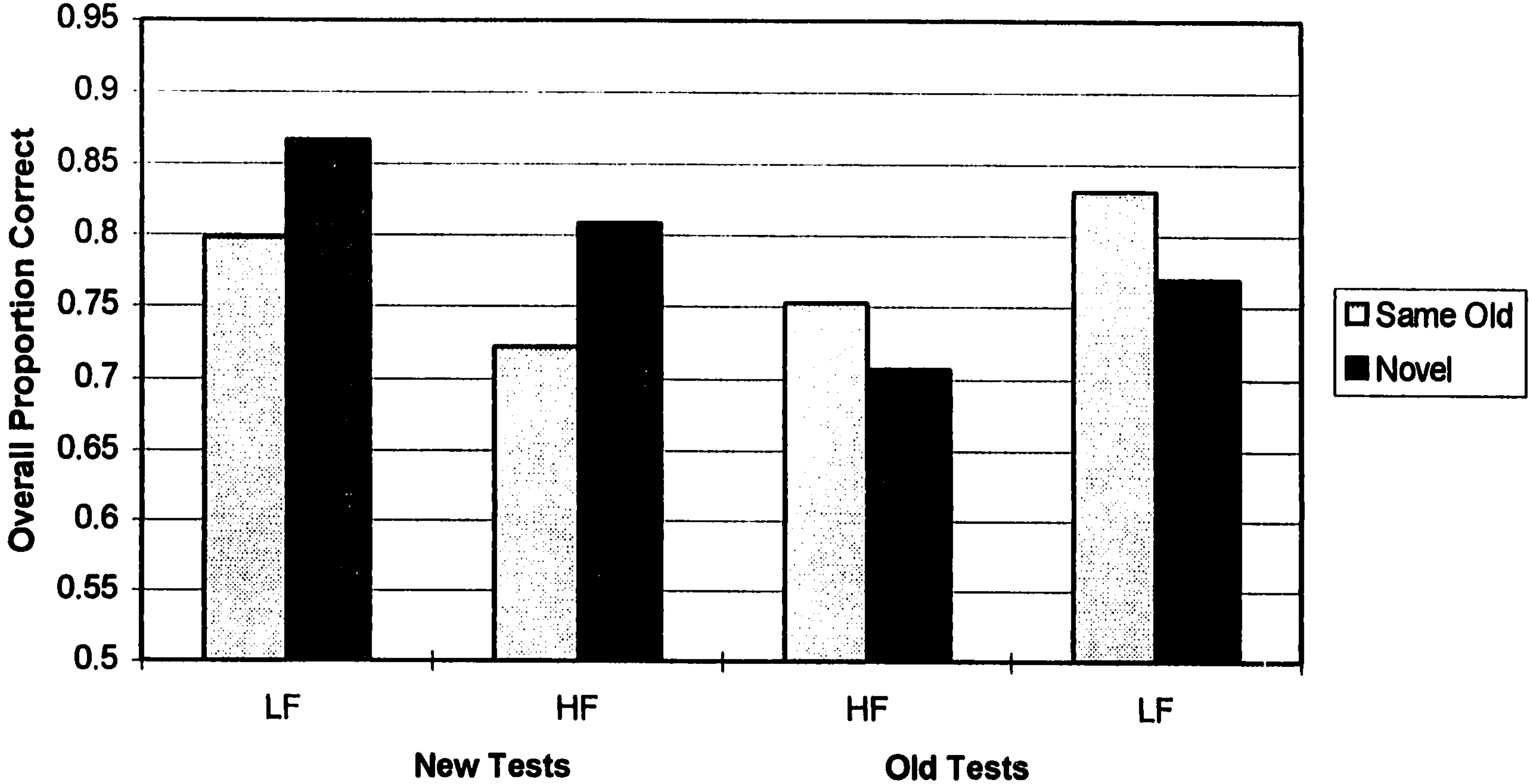

0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Overall Proportion Correct
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Novel

Figure 10

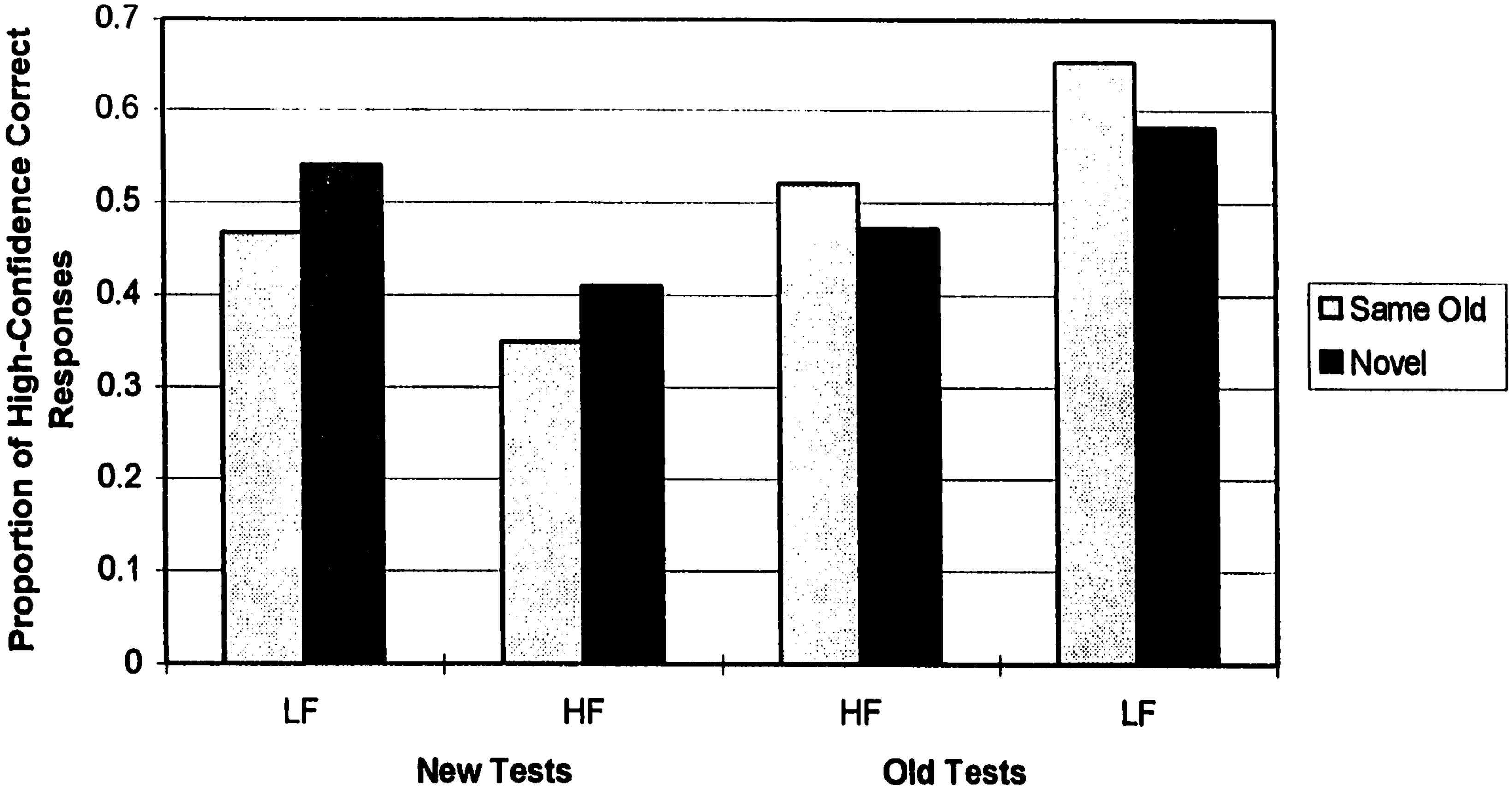

Proportion of High-Confidence Correct Responses
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
LF
HF
HF
LF
New Tests
Old Tests
Same Old
Novel

Figure 11

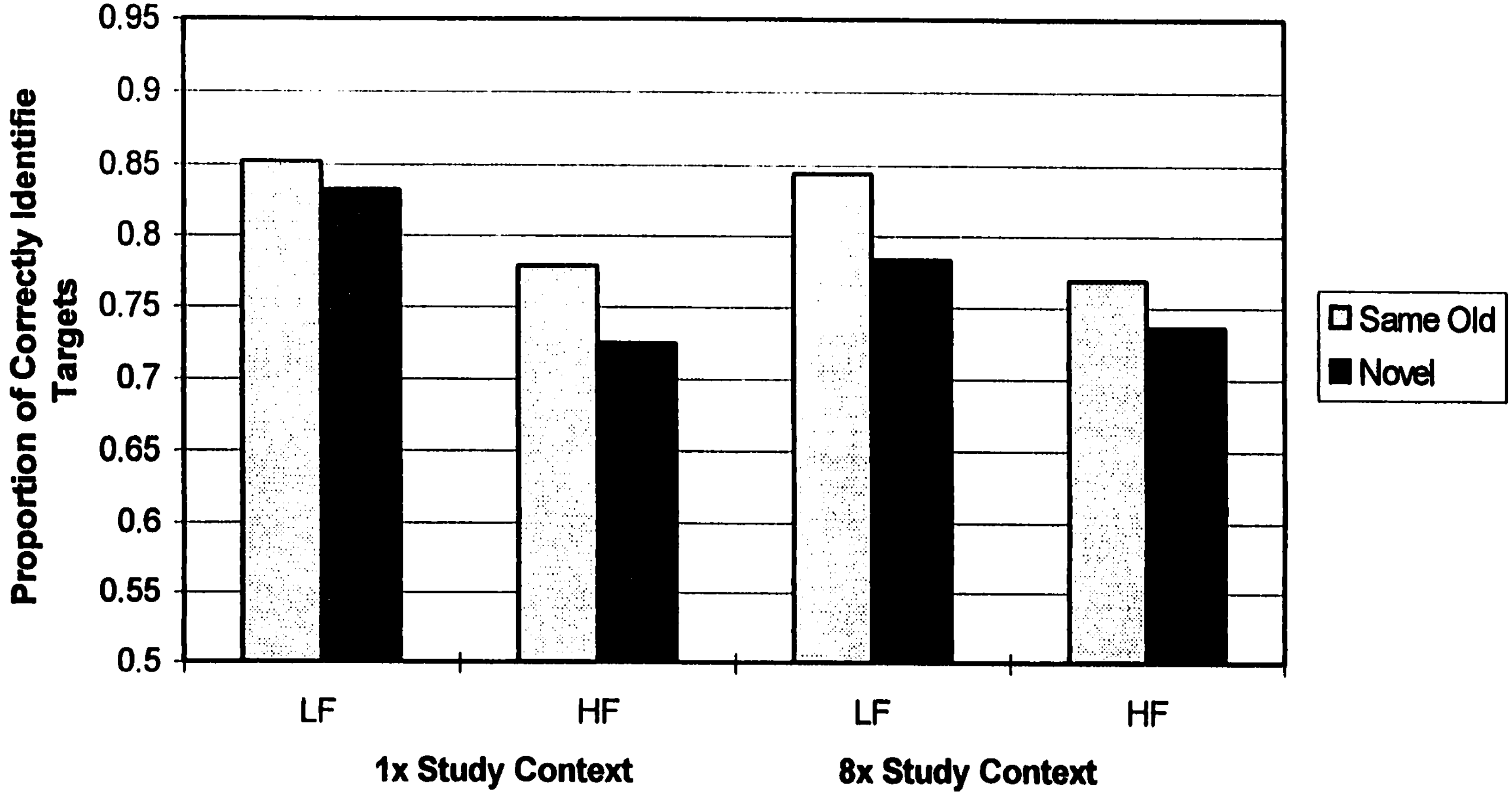

Proportion of Correctly Identifie Targets
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
LF
HF
LF
HF
1x Study Context
8x Study Context
Same Old
Novel

Figure 12

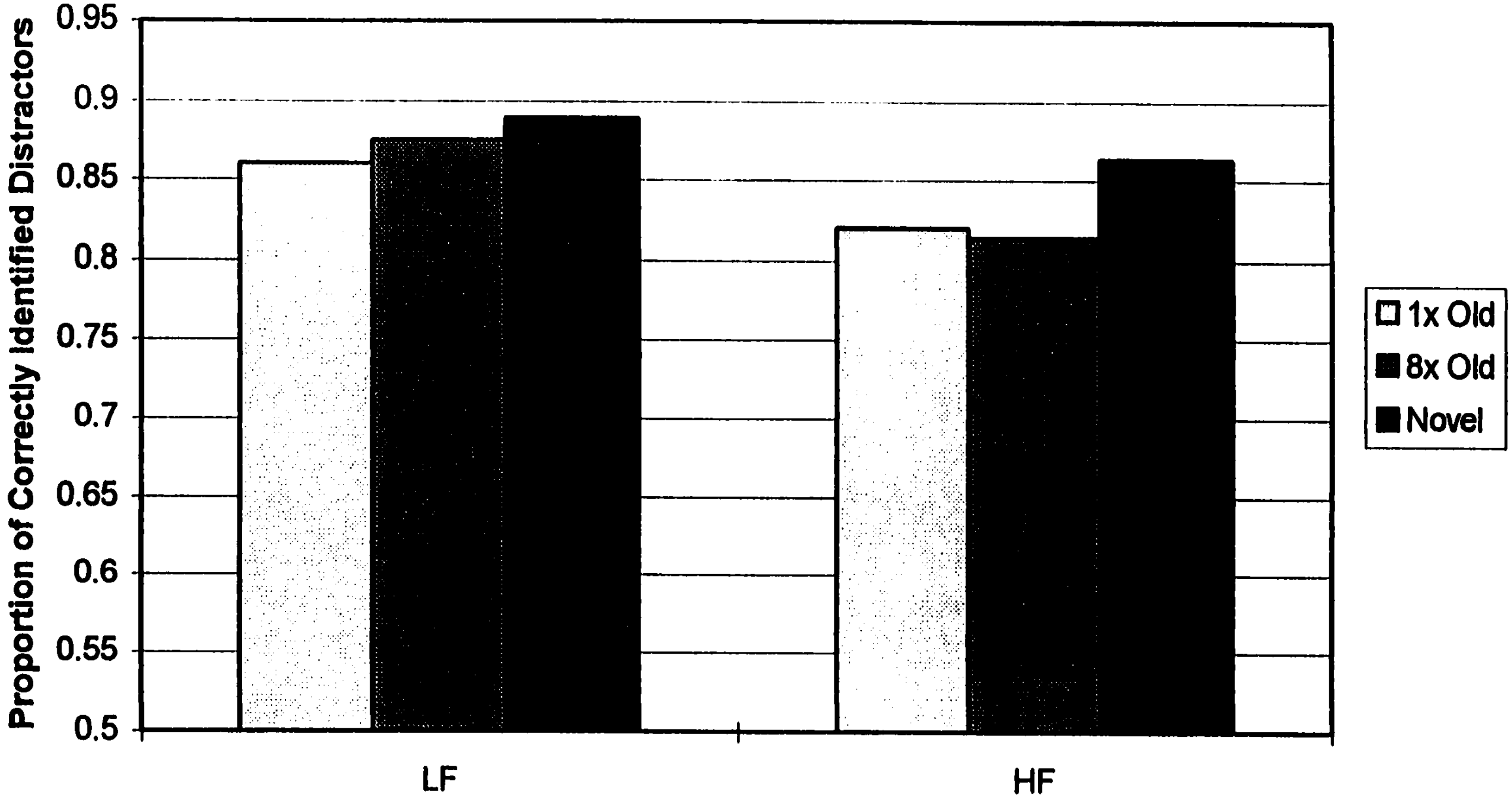
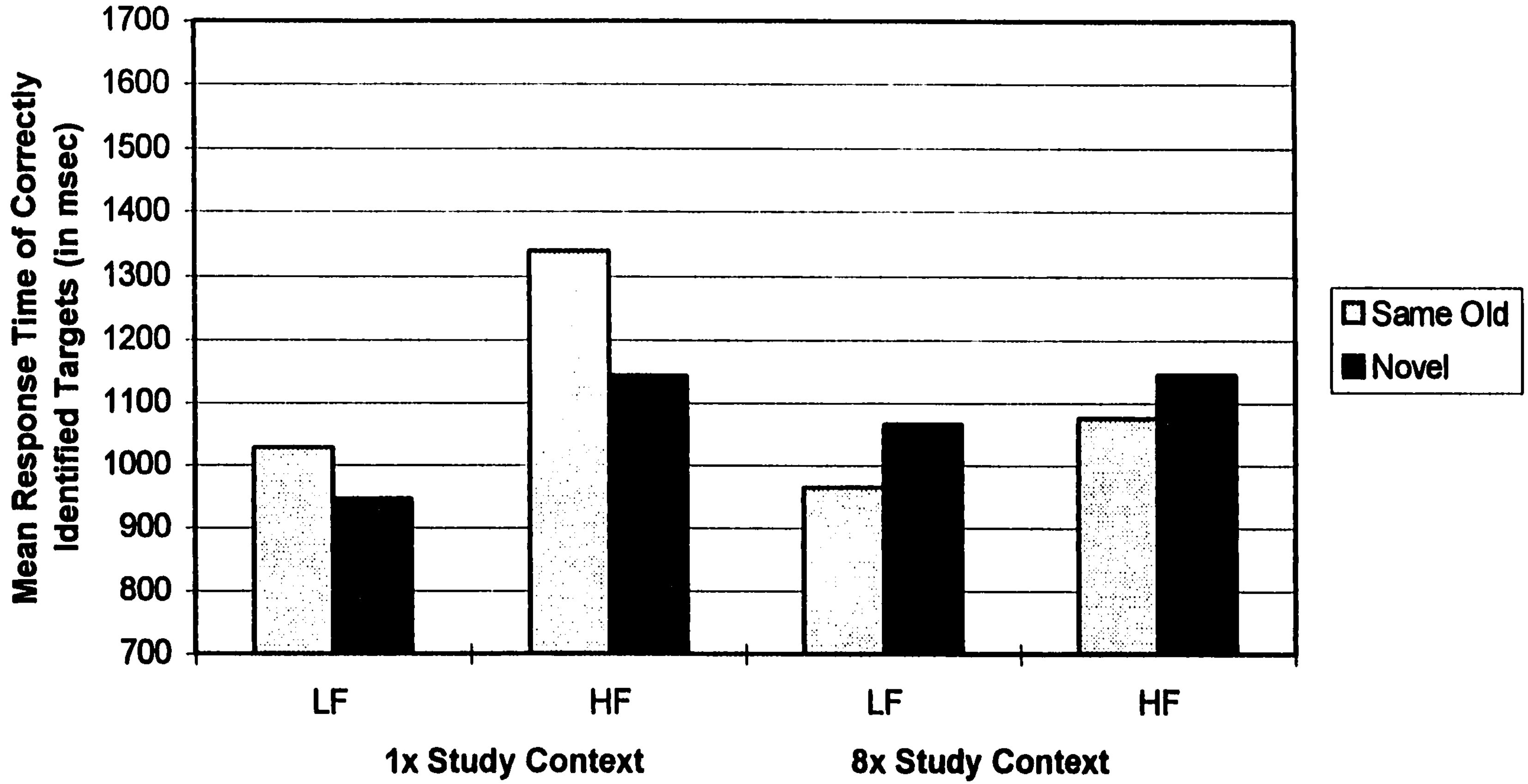

Proportion of Correctly Identified Distractors
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
LF
HF
1x Old
8x Old
Novel

Figure 13



Mean Response Time of Correctly Identified Targets (in msec)
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
LF
HF
LF
HF
1x Study Context
8x Study Context
Same Old
Novel

Figure 14

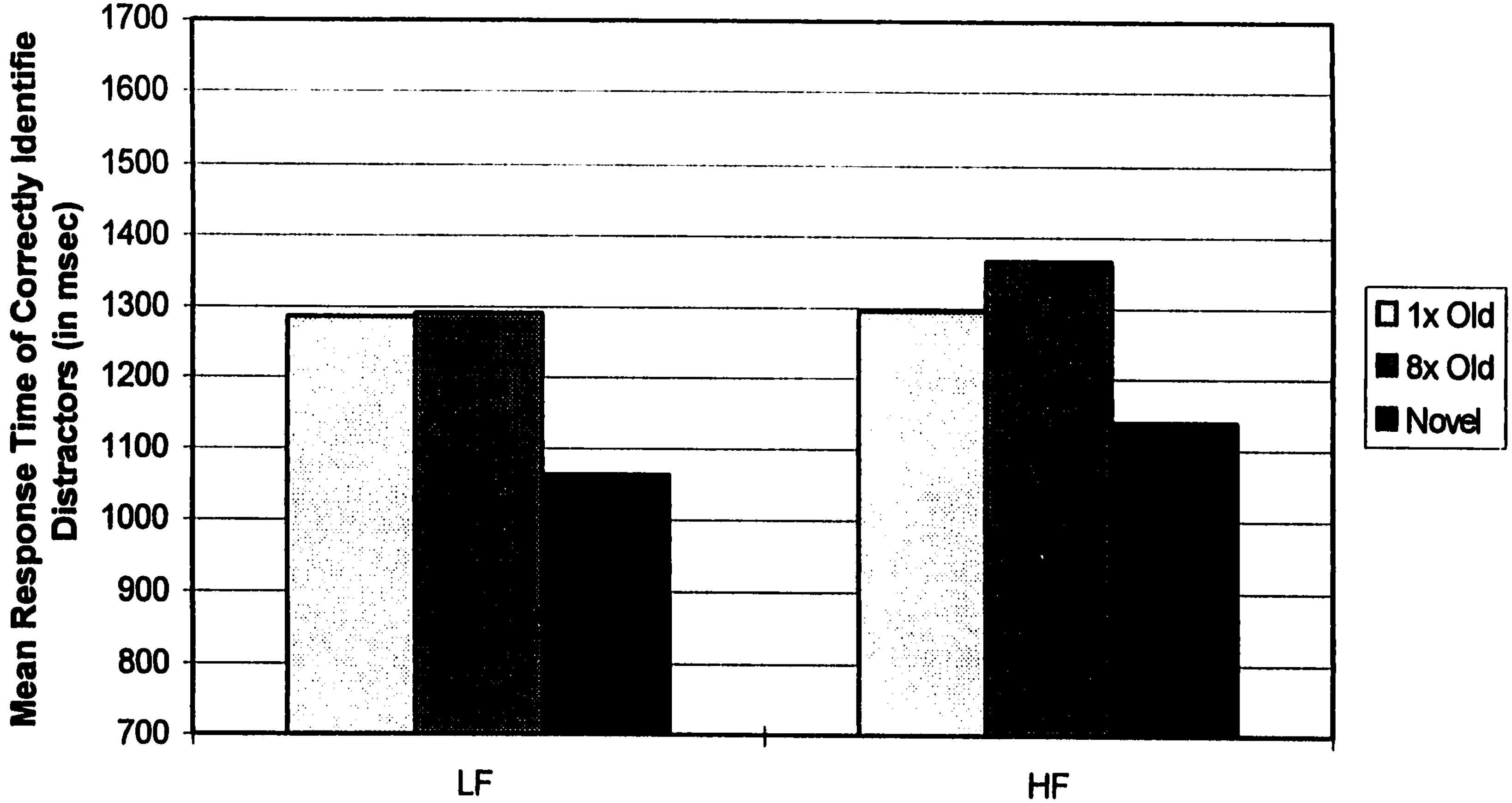

Mean Response Time of Correctly Identifie Distractors (in msec)
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
LF
HF
1x Old
8x Old
Novel

Figure 15

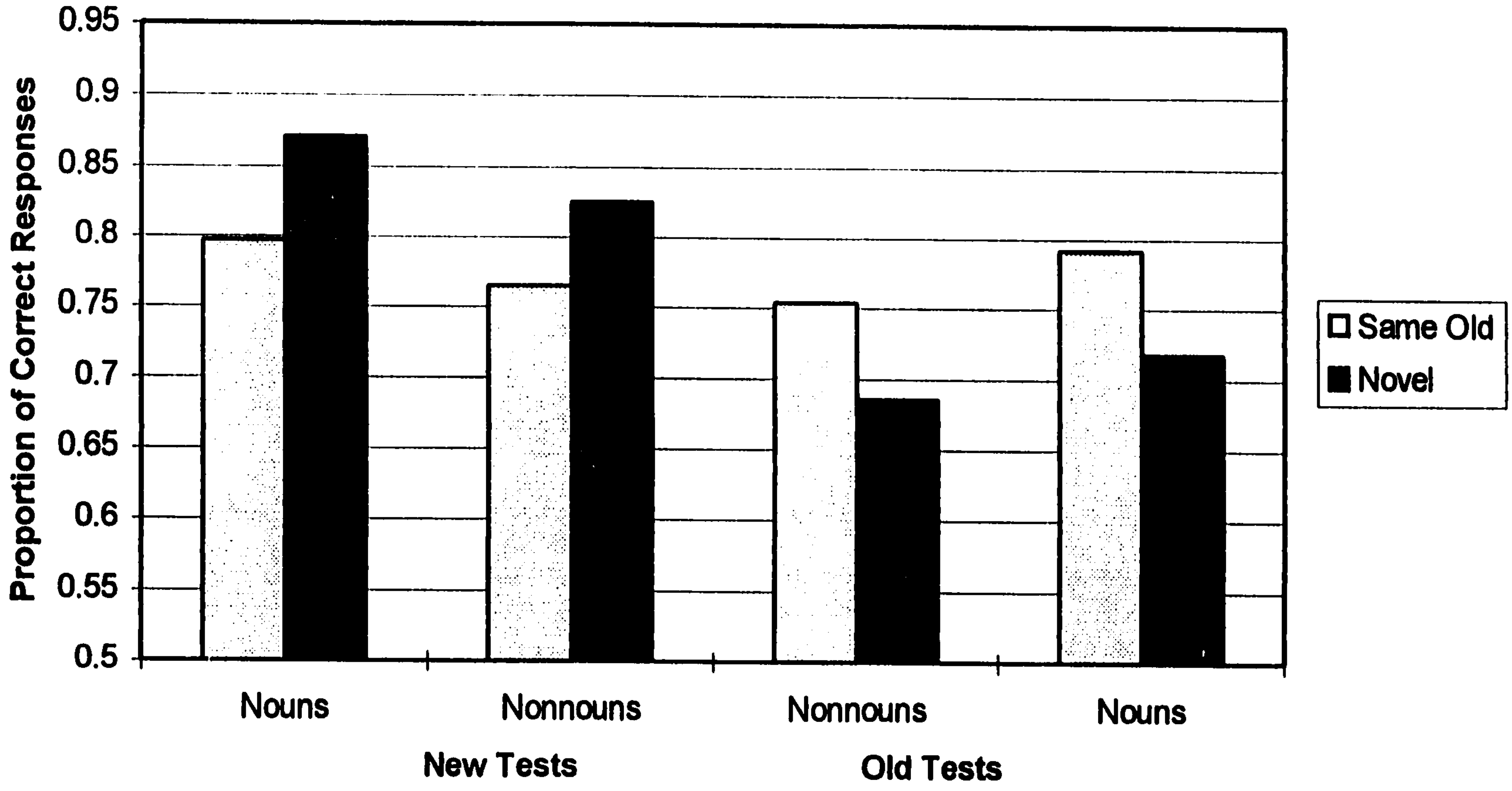

0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Proportion of Correct Responses
Nouns
Nonnouns
Nonnouns
Nouns
New Tests
Old Tests
Same Old
Novel

Figure 16

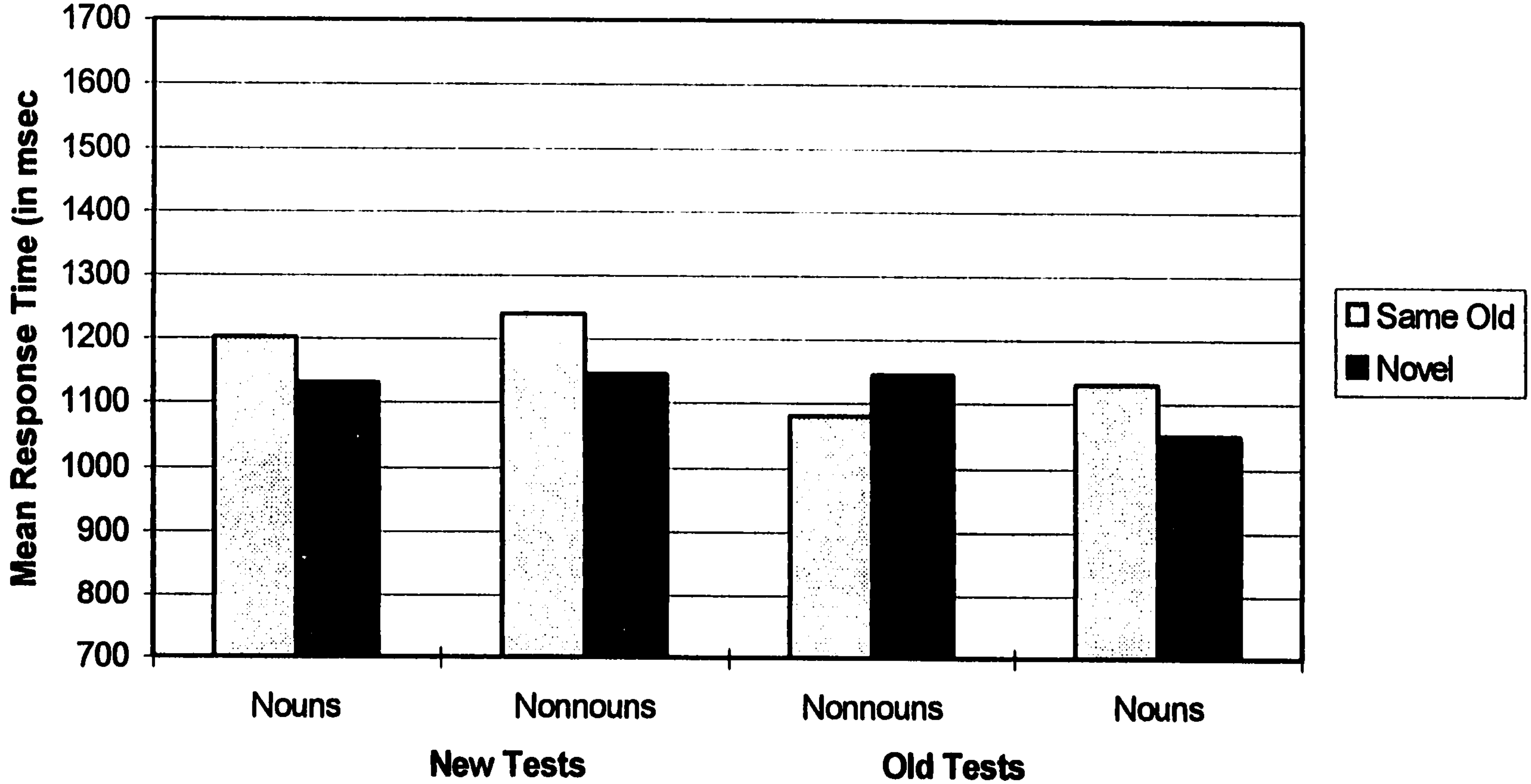

Mean Response Time (in msec
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
Nouns
Nonnouns
Nonnouns
Nouns
New Tests
Old Tests
Same Old
Novel

Figure 17

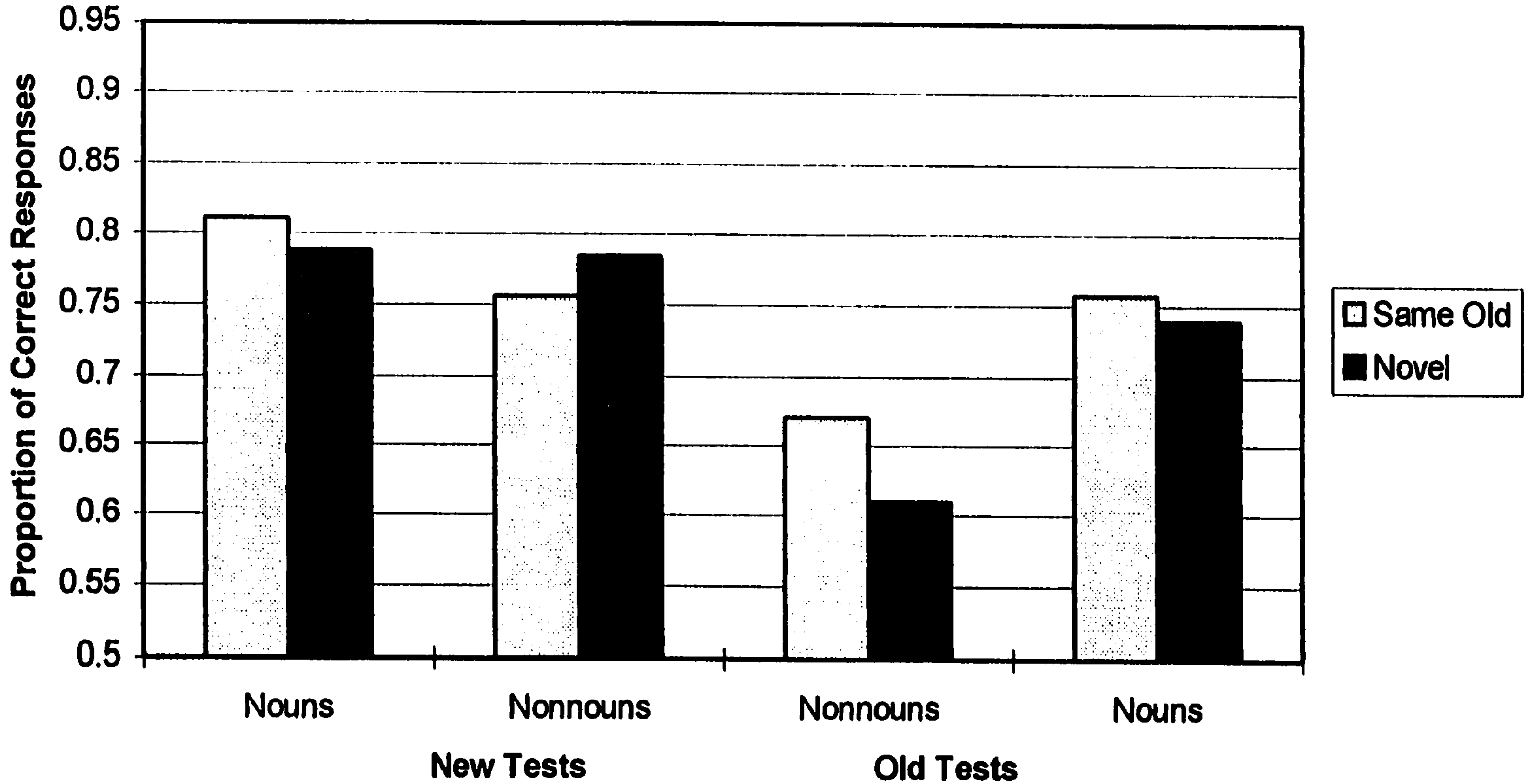
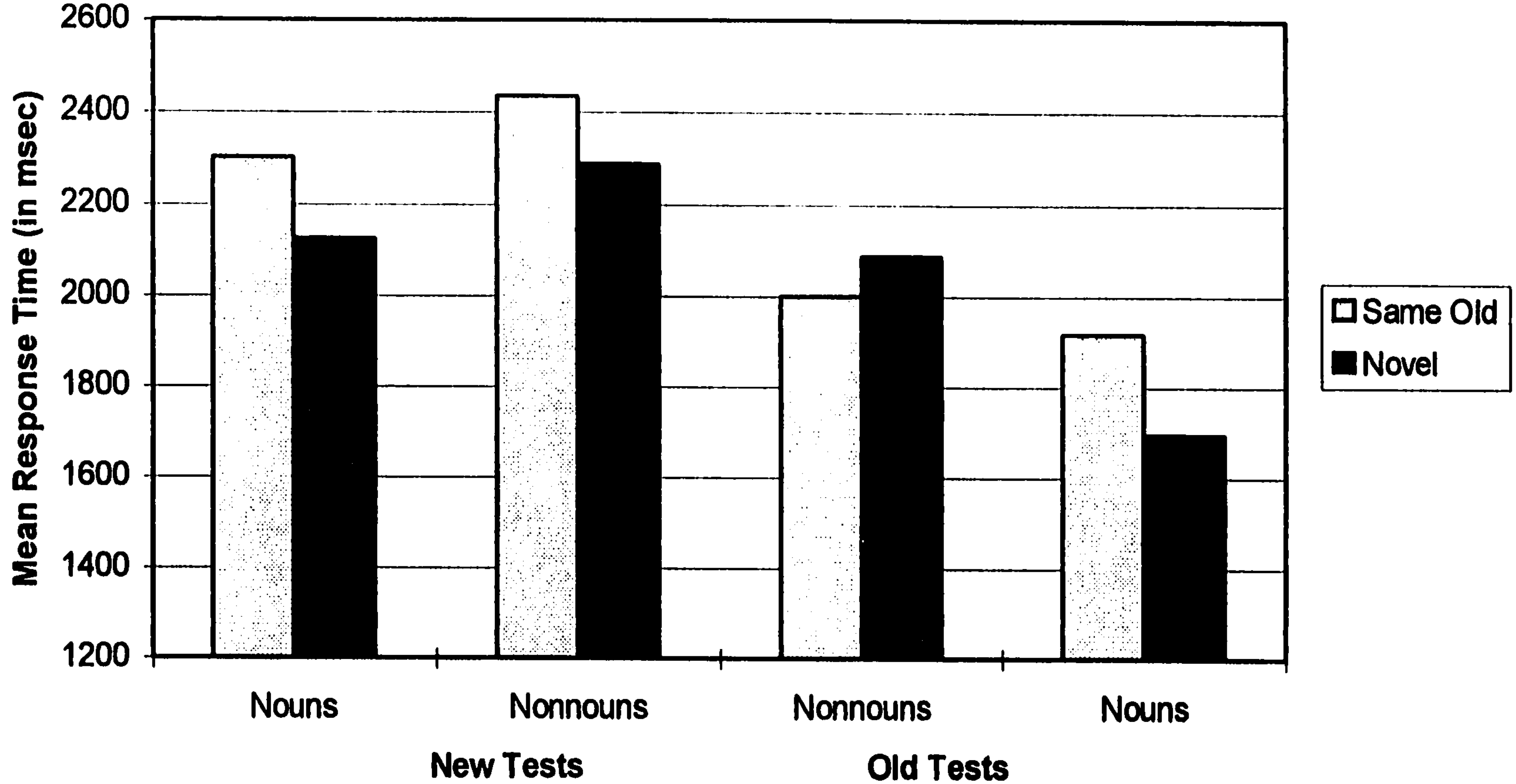

Proportion of Correct Responses
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Nouns
Nonnouns
Nonnouns
Nouns
New Tests
Old Tests
Same Old
Novel

Figure 18


Mean Response Time (in msec)
2600
2400
2200
2000
1800
1600
1400
1200
Nouns
Nonnouns
Nonnouns
Nouns
New Tests
Old Tests
Same Old
Novel

Appendix A

Instructions to Participants

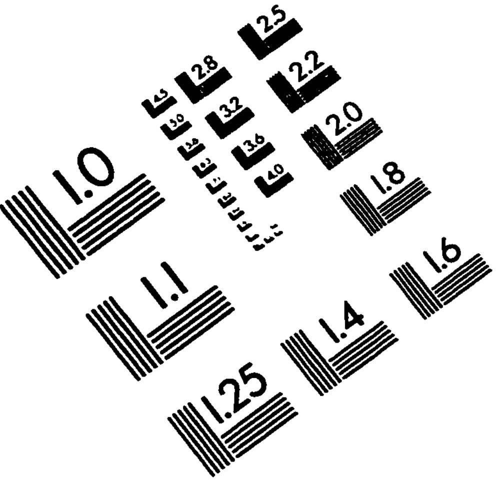
Thank-you for participating in this experiment. After seating yourself at the computer and making yourself comfortable, you will be able to begin the experiment whenever you are ready by pressing any key on the computer keyboard. When you do so you will start the first learning phase. During the learning phase you will be shown pairs of words, one pair at a time. The pairs of words will vary in colour and screen position. It will be helpful to try and associate these words together as this strategy can improve memory for the words. After all the word pairs have been shown the test phase will begin.

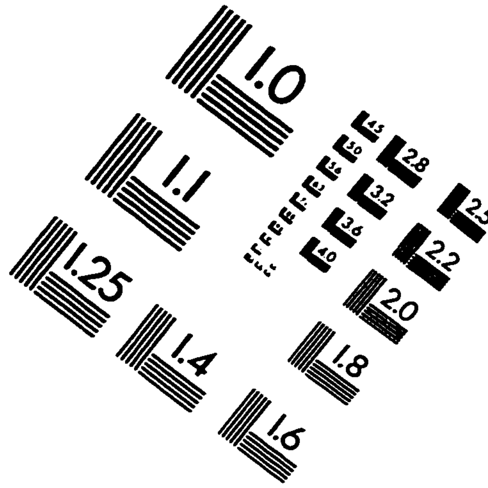
The test phase will consist of presentations of single words in varying colours and screen positions. When these words are presented, please indicate whether or not you recognize them from the word pairs you just learned. If you think the word was in the learning phase, press the key marked "old". If you do not think that the word was in the study phase, press the key marked "new". Please take a moment to orient yourself to which keys stand for "new" and "old" so as not to press the wrong key by mistake. Following the last presentation of the test phase, you will be able to take a break if you wish as the computer will wait for you to initiate the next learning phase by pressing any key on the keyboard. This will continue for 6 trials.

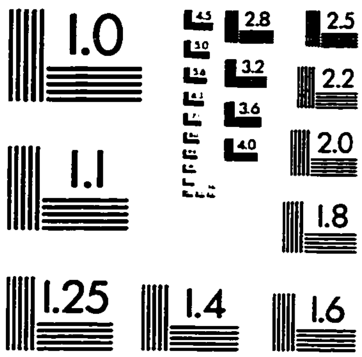
Following the last trial you will be informed that the session is complete. At this point please see the experimenter.

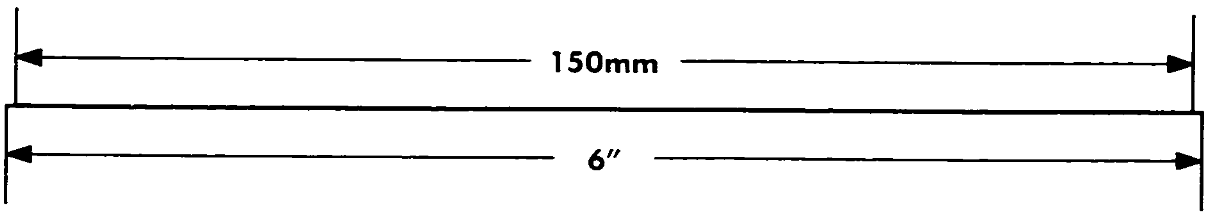
Again, thank-you for participating and if you have any questions about these instructions please ask the experimenter before you begin.

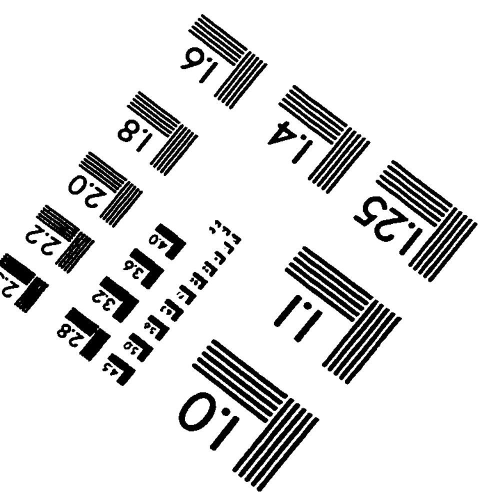
IMAGE EVALUATION TEST TARGET (QA-3)

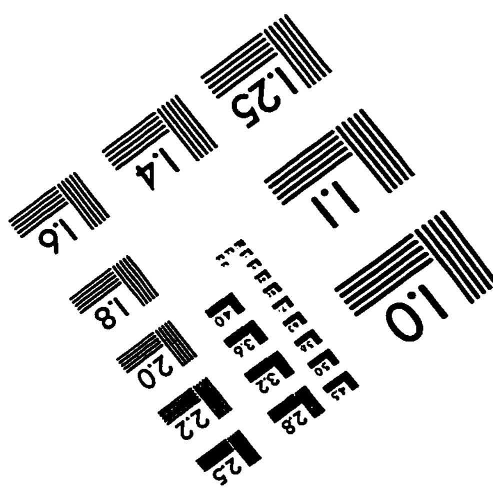












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989