The Colourful Experiment: Does Low Word Frequency and Tertiary Colours Affect Word Recognition?

Erinn Barry (131599830)

Dr. William Hockley

PP360 A

Wilfrid Laurier University

Abstract

Past research has suggested that the dual processing theory or the early-phase elevated attention hypothesis could explain the memory advantage observed in low frequency words for recognition memory. Our study examined the influence of local context (tertiary coloured words vs. black words) and word frequency (high vs. low) on word recognition in a Remember-Know procedure. We hypothesized that there would be an advantage in recollection for words that were both low frequency and tertiary coloured, given the fact that we assumed tertiary colours would further the distinctiveness of the word, and thus provide an advantage in recognition memory. We found a mirror effect for low frequency words, but we found that adding colour to words actually decreased the word frequency effect, which was contrary to our hypothesis. Consequently, we did not find full support for our hypothesis, nor were we able to fully explain our results with either the dual processing theory or the early-phase elevated attention hypothesis. However, we included suggestions for future research and possible reasons for why our hypothesis was not confirmed.

Keywords: word frequency, recognition memory, remember-know paradigm, colour

The Colourful Experiment: Does Low Word Frequency and Tertiary Colours Affect Word Recognition?

The purpose of this study was to investigate the effects of word frequency and local context on word recognition via the Remember-Know procedure. For this experiment, word frequency was either low or high frequency, and the local context was either tertiary coloured words or black words. The novel aspect of our experiment was whether tertiary coloured words combined with word frequency would influence recollection memory. We hypothesized that low frequency words in combination with tertiary coloured words would produce a greater recollection effect in the Remember-Know recognition procedure.

Previous researchers like Cook, Marsh, & Hicks, (2006) and Steyvers & Malmberg (2003) utilized the Remember-Know recognition procedure to examine word recognition for words with high or low word frequency. In the Remember-Know procedure, participants distinguished between making a response based on recollection or familiarity. If the participants identified an old response as being based on familiarity, there was an absence of recognition, similar to the feeling of déjà vu. In other words, the test item seemed familiar to the participants, but they could not identify any specific details about it. If the participants identified their old response as being based on recollection, this required participants to actually remember the details about a particular context, and those details aided the recollection of information. Consequently, in our experiment we utilized the Remember-Know word recognition procedure, whereby participants were required to make a judgment on whether a word was remembered or recollected, was familiar, or was a completely new word they had not seen during the study phase.

Cook et al. (2006) and Steyvers & Malmberg (2003) found evidence that the dual processing theory could help explain findings from the Remember-Know procedure. The researchers suggested that in the dual processing theory, words with low word frequency tended to create higher hit rates compared to words with high word frequency. The question was whether or not that difference in hit rate would be seen in the recollection responses or the familiarity responses. It was assumed that if there were higher false alarm rates in words with high frequency compared to words with low frequency, then that difference should be in the know responses, since know indicated familiarity. They found that words with low word frequency were recollected better, and that the difference in hit rates due to recollection and the difference in false alarm rates due to familiarity both supported the dual processing theory.

Many researchers have studied the effects of word frequency on item recognition performance (Steyvers & Malmberg, 2003; Cook, Marsh, & Hicks, 2006; Arndt & Reder, 2002). Steyvers & Malmberg (2003) suggested that the hit rate was greater and false alarm rate was lower for low frequency words compared to high frequency words. When this pattern of results occurred, it was called the mirror effect because the false alarm rate mirrored the hit rate. If the word was an old word the participants had seen before, then the hit rate was higher and the participants were more accurate in determining that the target was actually an old word. Similarly, if the word was a new word the participants had not seen before, they were less likely to say the word was old, which lowered the false alarm rate (Steyvers & Malmberg 2003). In other words, when word frequency was low there was an advantage for identifying old words when the words were indeed old, and there was also an advantage for rejecting new words when the words were actually new. Steyvers and Malmberg (2003) suggested words with low word frequency provided an advantage to the hit rate because those words were more distinctive, which made them easier to recollect during the test phase. Since low frequency words were more distinctive, there was less interference between items, and this resulted in better recognition memory (Reder et al., 2000).

Consequently, the dual processing theory could explain the low word frequency advantage found in recognition memory. This explanation proposed that low frequency words were easier to recollect, which produced a higher hit rate, and since those words were more distinctive and less familiar, they were immune to interference. Previous researchers (Steyvers & Malmberg, 2003; Reder, et al., 2000) postulated that words presented in fewer contexts were also lower in familiarity, which made those words easier to recollect, and this resulted in a lower false alarm rate because there were fewer competing contexts in memory. Indeed, Brown, Lewis, and Monk (1977) suggested that since low frequency words were encountered less frequently, they must be inherently more distinctive than high frequency words. In terms of the resulting mirror effect observed in the word frequency effect, the dual processing theorists posited that participants recollected low frequency words more often than high frequency words, and this created the hit rate in the mirror effect (Arndt & Reder, 2002). Furthermore, it was theorized that when participants found high frequency words more familiar than low frequency words, this difference created the false alarm rate found in the mirror effect (Arndt & Reder, 2002).

Additionally, Arndt and Reder (2002) suggested measurements of recollection should be greater for low frequency items compared to high frequency items, which was precisely what we hypothesized for our experiment. The dual processing theory suggested that the hit rate advantage for low frequency items should be due to differences in recollection, so there should be greater levels of recollection for low frequency words compared to high frequency words (Arndt & Reder, 2002). Similarly, Reder et al. (2000) proposed that high frequency words would prompt more familiar responses because high frequency words were less likely to produce a remember response for both hits and false alarm rates.

Steyvers and Malmberg (2003) suggested that false alarm rates in studies of recognition memory had to be based on familiarity and not recollection, since there were no words to recollect. They found that words that were high in frequency appeared to be more familiar, which resulted in a lower hit rate. Words that had higher word frequency also had higher rates of familiarity because people encountered these types of words more regularly in different contexts during the pre-experiment phase. Similarly, low frequency words had less familiarity associated with them, so they were more distinctive, and this resulted in better recollection.

Malmberg and Nelson (2003) and Arndt and Reder (2002) both studied the effects of low frequency words in word recognition tests. Their results indicated that for recognition tests, low frequency words tended to be better recalled than high frequency words. One possible answer for this effect may be due to the fact that recognition tests provided more cues for participants' memory, which potentially made it easier for participants to create associations with low frequency words.

Malmberg and Nelson (2003) posited the elevated attention hypothesis as an alternative to the dual processing theory for explaining the low frequency advantage seen in recognition memory. They suggested that low frequency words were harder to process, so participants tended to devote more attention to them, and this additional attention produced the advantage seen in recognition tests. Based on their results, Malmberg and Nelson (2003), argued that the elevated attention hypothesis was really an early phase elevated attention hypothesis. This hypothesis suggested that the processing of low frequency words must have happened early on, and lower frequency words were more structurally difficult to read, so participants were required to devote more resources to reading a low frequency word compared to a high frequency word. As such, low frequency words received an advantage because of the extra attention required during the early reading stage. More recently, Criss and Malmberg (2008) showed agreement with the early-phase elevated attention hypothesis, such that low frequency words were comprised of uncommon features which made their identification more difficult, and this difficulty during the early-phase resulted in the low frequency advantage.

Arndt and Reder (2003) also studied recognition memory. However, in this study, Arndt and Reder (2003) incorporated unusual fonts in their stimuli. The researchers included three different conditions. In the unique font condition, each of the 72 study items were presented in a different font. In the correlated condition, six fonts were correlated with the semantic themes of the test items, so all of the study items from a particular theme were associated with one particular font. In the standard condition, all of the study items were presented in one common font, which was Helvetica in this case. Their results suggested that participants had greater recognition when the unusual fonts were used in the unique and correlated conditions, but not for the standard condition (Arndt & Reder, 2003). Consequently, these findings were the basis for why we modified the distinctive visual properties of our stimuli to include either tertiary colours or black text, rather than differing fonts.

Given the fact that Cook et al. (2006) and Steyvers & Malmberg (2003) utilized the Remember-Know recognition procedure, Steyvers and Malmberg (2003) studied the effects of word frequency on item recognition, Arndt and Reder (2002), Reder at al. (2000), Cook et al. (2006) and Steyvers & Malmberg (2003) posited the dual processing theory as an explanation for the word frequency effect, Malmberg and Nelson (2003) and Criss and Malmberg (2008) proposed the early-phase elevated attention hypothesis as an explanation for the word frequency effect, and the fact that Arndt and Reder (2003) studied recognition memory by incorporating unusual fonts as their stimuli, the rationale behind our current study was whether there would be an effect of word frequency in combination with local context in the Remember-Know procedure. Consequently, we wondered whether tertiary coloured words in combination with low frequency words would be recollected better because tertiary colours were hypothesized to be more distinctive, akin to low frequency words. Our hypothesis was that there would be an advantage in recollected responses in our Remember-Know procedure if the word was low in frequency, but there should be a further advantage in recollection if the word was both low in frequency and was tertiary coloured.

Method

This experiment was a 2 (test: old vs. new) x 2 (word frequency: high vs. low) x 2 (local context: tertiary vs. black) repeated measures design. There were 3 dependent variables. The first dependent variable was a measure of remembered or recollected responses. The second dependent variable was a measure of familiar or know responses. The last dependent variable was the total old responses collapsed over remember and familiar responses. The aim of this experiment was to investigate the impact of local context and word frequency on word recognition within the Remember-Know procedure.

Participants

The participants in this study were registered undergraduate students attending a third year laboratory research course in cognitive psychology at Wilfrid Laurier University. There were a total of 18 male and female participants, and no other demographic information was collected. It is important to note that two participants misunderstood the instructions that were given, so their data was removed from the analysis.

Apparatus and Materials

For this study, 50 low word frequency (LWF) and 50 high word frequency (HWF) words were selected from the Touchstone Applied Science Associates, Inc. (TASA) corpus database, for a total of 100 words, as recommended by Meeks, Knight, Brewer, Cook, and Marsh (2014). All of the words were common nouns, and contained three to eleven letters. Furthermore, the words in this experiment were between 10-2020 on the word frequency scale, which was supplied by the TASA corpus database. To be more specific, the low frequency words were between 10-50 on the word frequency scale, and the high frequency words were between 2020-27491. The stimuli were presented using PowerPoint software and a screen projector at the front of the classroom. All of the test items were in Times New Roman size 48 against a white background in the centre of the screen. In addition, the test items were presented either in black font or in one of three tertiary colours: green-yellow, blue-purple, or red-orange.

Procedure

In terms of the procedure for this experiment, it was a single experimental session that consisted of one study list and one test list. The study list contained 52 words that were chosen at random. The first and last two words in the study list were not tested in order to avoid the possibility of primacy and recency effects. The study words were either LWF or HWF, and they were either coloured in one of three tertiary colours (green-yellow, blue-purple, or red-orange), or they were in black font. Therefore, the study list contained 12 LWF/tertiary words, 12 LWF/black words, 12 HWF/tertiary words, and 12 HWF/black words, for a total of 48 words. The buffer words consisted of two LWF/tertiary words, one HWF/tertiary word, and one HWF/black word. For the study list, each word was presented for 3s. The order of the study words was chosen at random with the constraint that each word and its corresponding colour and level of word frequency be presented equally in each half of the list, and only two of the same frequency or coloured words could appear successively on the list. The participants were told to only memorize the words that were presented, and that memory of the colour of the words would not be tested.

The test list was comprised of 96 words that were presented in random order for 6s. The 48 words that were used previously for the study list were included, as well as 48 new study list items. As before, the 48 new study list items consisted of 24 LWF words with 12 words in black font and 12 words in tertiary font, and 24 HWF words with 12 words in black font and 12 words in tertiary font. After viewing each test item, participants were instructed to circle an "Old-Remember" response, an "Old-Familiar" response, or a "New" response in order to measure recognition memory. Our instructions to the participants for when they should circle remember or when they should circle familiar were taken from Gardiner and Java's explanation of the Remember-Know procedure (1990). The participants then circled the corresponding answer on the response sheet. The retention interval between the study list and the test list was approximately 2 min, whereby the test instructions and response sheets were given to the participants during this time.

This study was designed to examine the effects of word frequency and tertiary colours on word recognition via the Remember-Know procedure. The unique feature of this study was whether the effects of word frequency and tertiary colours influenced participants' ability to remember the old study items. Given the previous literature, we hypothesized that the combination of low frequency words and tertiary colours should facilitate recognition further compared to to low word frequency and tertiary colours alone. Our three independent variables were old versus new test items, high frequency versus low frequency words, and colour versus black for local context. We scored the old responses in order to measure hits and false alarms. In addition, we scored the old remember responses, old familiar responses, and the total old responses collapsed over recollection and familiarity.

Results

For this experiment, we only scored the old responses. Therefore, our three dependent variables were recollection, familiarity, and a measure of overall old responses collapsed over recollection and familiarity. In order to establish whether an effect was significant, the main effects and potential interaction were investigated using the conventional alpha level of .05 that was chosen ahead of time. Consequently, we completed three different analyses.

In the first analysis, we sought to determine the effects of colour and word frequency on word recollection. Therefore, we examined the main effects of test (old vs. new), word frequency (high vs. low) and context (tertiary vs. black) in a repeated measures analysis of variance (ANOVA). Since we only scored the old test items, higher mean scores for hit rates indicated better performance. The main effect of context was not significant: $F(1, 15) < 1$, ns. The main effect of word frequency was also not significant: $F(1, 15) = 1.078$, $MSE = 2.348$, $p = .316$. The main effect of test was significant: $F(1, 15) = 67.470$, $MSE = 6.781$, $p < .001$. Consequently, for this main effect of test, the recollection responses for old items ($M = 4.984$, $SE = .513$) was significantly higher than the recollection responses for new items ($M = 1.203$, $SE = .396$). This result indicated that memory for old test items was above chance for recollection responses.

Four interactions were analyzed. For the first interaction we measured context x word frequency, but it was not significant: $F(1, 15) = 1.697$, $MSE = 1.842$, $p = .212$. For the second interaction we measured context x test, and it was also not significant: $F(1, 15) = 1.392$, MSE =

.808, $p = .256$. Interestingly, the word frequency x test was significant: $F(1, 15) = 29.917$, $MSE = 1.508$, $p < .001$. For this interaction, there was a larger effect of word frequency when the target word was either low or high frequency (see Table 1). The last interaction of context x test x word frequency was not significant: $F(1, 15) < 1$, ns.

In the second analysis, we sought to determine the effect of local context and word frequency for word familiarity. As such, we examined the main effects of test (old vs. new), word frequency (high vs. low) and local context (tertiary vs. black) in a repeated measures ANOVA. Once again, a higher mean score indicated a better hit rate. The main effect of context was not significant: $F(1, 15) < 1$, ns. The main effect of word frequency was significant: $F(1, 15) = 9.765$, $MSE = 4.381$, $p = .007$. For this effect, the mean for the familiarity responses for high frequency words ($M = 4.406$, $SE = .393$) was higher than the mean for the familiarity responses for low frequency words ($M = 3.25$, $SE = .312$). This result suggested that participants were more likely to indicate familiarity for a high frequency word compared to a low frequency word. We also measured the main effect of test for word familiarity, but it was not significant: $F(1, 15) < 1$, ns.

Four interactions were analyzed. Context x word frequency on word familiarity was not significant: $F(1,15) < 1$, ns. Context x test was not significant: $F(1, 15) < 1$, ns. Word frequency x test was not significant: $F(1, 15) = 2.276$, $MSE = 2.692$, $p = .152$. Although this last interaction was not significant, the result indicated a trend toward word frequency, such that participants were more likely to say familiar for a new word if that word was a high frequency word ($M = 4.5$, $SE = .461$) compared to a new word that was low frequency ($M = 2.91$, $SE = .377$). Furthermore, if it was an old word, participants were just as likely to say familiar whether or not the word was a high frequency word ($M = 4.31$, SE .526) or if the word was a low frequency word (3.6, $SE = .411$). In other words, familiarity changed as a function of word frequency for target items. The last interaction was context x test x word frequency and it was not significant: $F(1, 15) < 1$, n.s.

For the last analysis, we wanted to determine the effects of colour and word frequency on the overall old responses collapsed over familiarity and recollection. Therefore, we examined the main effects of test (old vs. new), word frequency (high vs. low), and local context (tertiary vs. black) in a repeated measures ANOVA. The main effect of context was not significant: $F(1, 15) < 1$, ns. The main effect of word frequency was significant: $F(1, 15) = 13.915$, $MSE = 4.249$, $p = .002$. For this main effect of word frequency, the mean for high frequency words ($M = 7.56$, $SE = .50$) was higher than the mean for low frequency words ($M = 6.20$, $SE = .471$). This result suggested that for overall old responses, collapsed over recollection and familiarity, participants were more likely to indicate an old response for a word with high word frequency compared to a word with low word frequency. The main effect of test was significant: $F(1, 15) = 93.992$, $MSE = 5.320$, $p < .001$. In this case, the mean for old test items ($M = 8.86$, $SE = .455$) was higher than the mean for new test items ($M = 4.90$, $SE = .531$). Therefore, participants were more likely to choose an old response when the item was really an old item, compared to when the item was actually a new item.

We analyzed four interactions. The first interaction was between context x word frequency, and it was significant: $F(1, 15) = 6.981$, $MSE = 4.883$, $p = .018$. In terms of the effect of high word frequency for this interaction, the mean was larger for the black condition ($M = 7.813$, $SE = .547$) compared to the coloured condition ($M = 7.313$, $SE = .518$). Alternatively, in terms of the effect of low word frequency on this interaction, the mean was only slightly larger for the colour condition ($M = 6.344$, $SE = .518$) compared to the black condition ($ME = 6.063$, SE = .472). The interaction between context and test was not significant: $F(1, 15) < 1$, ns. The interaction between word frequency and test was significant: $F(1, 15) = 40.884$, $MSE = 2.270$, $p < .001$. For this interaction, the mean for a target response with low word frequency ($M = 9.03$) was larger than the mean for a distractor response with low word frequency ($M = 3.38$). Similarly, the mean of a target response with high word frequency ($M = 8.67$) was larger than the mean of a distractor response with high word frequency ($M = 6.438$). In other words, participants' overall recognition memory was above chance because there was a large effect of test. The last interaction was context x test x word frequency, and it was not significant: $F(1, 15) < 1$, ns.

Discussion

This study aimed to explore the effects of word frequency and tertiary colours on item recognition via the utilization of the Remember-Know procedure. We hypothesized that low frequency words would have an advantage for recollected responses compared to high frequency words, and that the addition of tertiary colours would provide a further recollection advantage for low frequency words. To test word frequency, we varied the words, such that half of the stimuli were low frequency words and the other half were high frequency words. Additionally, in order to test local context, we varied the colour of the stimuli, such that each stimulus item appeared as green-yellow, blue-purple, red-orange, or in a black font.

Our results indicated that memory was above chance for recollection. Since this effect was not present for familiarity, participants seemed to be treating their familiarity judgments like confidence judgments, meaning they were basing the familiarity judgment on whether the word seemed familiar. Whereas for recollection, participants claimed to remember the study item, so recollection was not treated like a confidence judgment. There was also an interaction between level of word frequency and whether the item was a target or a distractor. Consequently, we found a mirror effect, whereby low frequency words had a higher hit rate for targets, and as such were better recollected in comparison to high frequency words (see Table 4). The mirror effect we observed replicated previous findings from Arndt & Reder (2002), Steyvers & Malmberg (2003), and Cook et al. (2006), whereby low frequency words had a higher hit rate and lower false alarm rate compared to high frequency words. Interestingly enough, we did not find this mirror effect for familiarity. Therefore, for recollection responses there was a larger effect when the word was a target word compared to a distractor word, and as a result, when a target word had low word frequency, participants made more claims of recollection for that item.

Our results supported the finding that words with low frequency provided an advantage to the hit rate because those words were more distinctive (Steyvers & Malmberg, 2003). Reder et al., (2000) also proposed low frequency words were more distinctive in comparison to high frequency words, and because of this distinctiveness, there was less interference between items, which resulted in better recognition memory. Steyvers and Malmberg (2003), and Reder et al. (2000) suggested the dual processing theory could explain the low word frequency advantage in recognition memory. Our results supported this conclusion, given the fact that in our experiment low frequency words were claimed to be recollected more than high frequency words when we measured recollection. We also replicated the mirror effect for recollection, whereby participants claimed recollection for low frequency words more often than high frequency words, as posited by Arndt & Reder (2002).

In our experiment, familiarity responses for high frequency words had a higher mean compared to familiarity responses for low frequency words. This finding supported a prediction by Reder et al. (2000), whereby high frequency words should elicit greater familiar responses, since high frequency words were less likely to produce a remember response for targets and distractors, and this was exactly what we found. We also found support for Steyvers and Malmberg's (2003) prediction that words that were high in frequency were more familiar, which resulted in a lower hit rate for recollection. In our experiment, when participants indicated old for a false alarm, this was another way of guessing, and this guess was based on the familiarity of the test item, as evident by the fact that our main effect of test for familiarity was not significant.

This was clear evidence that when participants were lacking recollection memory for old versus new items, they were using the familiarity of the words to influence their decision. Therefore, if a word was a high frequency word, and that level of frequency made the word inherently more familiar, this familiarity made the participant more likely to choose it, which supported findings by Reder et al. (2000), Steyvers and Malmberg (2003), and Arndt & Reder (2002). Furthermore, our participants showed a clear difference in discrimination between items that were recollected or remembered compared to items that were familiar. When participants were unable to discriminate, there was a clear effect of familiarity. In terms of familiarity, regardless of whether the test item was a target or a distractor, high frequency words had the larger mean in comparison to low frequency words.

The interaction of word frequency by test for familiarity was not significant. However, since we only had a small number of participants in our experiment, we potentially lacked power, and given this consequence, it is fair to discuss this interaction as a trend. Therefore, our results indicated that word frequency was larger for distractor items, so participants were more likely to say familiar to a target word despite the level of word frequency. This trend was towards word frequency affecting targets, which would have created the mirror effect if it had been significant. As shown in Table 2, the familiarity for low frequency distractors was lower than the mean for the familiarity for targets. This meant that participants were more likely to say a word was familiar if the word was low frequency and it was a target. Conversely, if the word was a high frequency word, participants were more likely to say a word was familiar if it was a target or a distractor word. In other words, we were able to replicate one element of the mirror effect, since we found a mirror effect for low frequency words, but not for high frequency words, but this was only a partial replication.

In terms of our results for the overall old responses, there was a main effect of word frequency. Given that we also found a main effect of word frequency for familiarity but not for recollection, and our measure of old responses was collapsed over familiarity and recollection, it was likely that the familiarity responses were affecting word frequency for this overall measure, so familiarity was driving this effect. Nevertheless, there was a main effect of test for recollection, which showed that memory was above chance. In this case, recollection was driving this effect because there was no main effect of test for familiarity. Therefore, word recognition was above chance, and the participants had tried to remember the study items they were supposed to remember.

There was an interaction between context and word frequency for overall old responses. This was due to the fact that the effect of word frequency was larger in the black condition compared to the coloured condition. In other words, there was a larger mean for high frequency words in the black condition, but high and low word frequency was almost equal for the coloured condition. By taking the difference between word frequency and each level of context for targets and distractors, we were able to examine this effect more clearly. First, we found the difference between HWF/tertiary (M = 7.313) and LWF/tertiary (M = 6.344), which was 1.0. Second, we found the difference between HWF/black (M = 7.813) and HWF/tertiary (M = 6.063), which was

1.7. This difference meant that by adding colour to a high frequency word, we actually lowered the probability that participants would say old to an old item. Or, said differently, by adding colour to a word we actually decreased the effects of word frequency.

Interestingly, we found another mirror effect for word frequency in the overall responses (see Table 5). There was a large effect of test, so memory was again above chance for saying an old test item was old when it was really old, and rejecting a new item when it really was new. Given that the word frequency x test interaction was not significant for familiarity, although there was a trend in that direction, recollection was the factor that was driving this mirror effect.

Since there was a difference in the pattern of results between recollection and familiarity, we interpreted these results as support for the utilization of the Remember-Know procedure for distinguishing between recollection and familiarity. For example, when the participants claimed to remember a test item they circled the Old-Remember response, which suggested they were making a distinction between recollection and familiarity, and their ability to discriminate was very different depending on whether they were claiming a remembered response or whether they were just making a familiarity judgment. As such, there was a different pattern of results for whether participants responded based on recollection or familiarity. The effects of word frequency emerged in the participants' familiarity judgments because they based their recognition decision on the familiarity of the word. When the word was high frequency it was more familiar, so the participants were more likely to make a familiarity judgment.

Ultimately, our results suggested that memory was above chance for claims of recollection, participants were treating their familiarity judgments like confidence judgments, and there was a mirror effect for recollection, whereby low frequency words had a higher hit rate than high frequency words. Additionally, when participants could not discriminate between items that were recollected versus items that were familiar, there was a clear effect of familiarity, because high frequency words, which were inherently more familiar, were chosen more often than low frequency words. Although our interaction between word frequency and test for familiarity was not significant, there was a trend toward a mirror effect. In terms of the overall old responses, there was an interaction between context and word frequency, whereby the effect of word frequency was larger in the black condition. By taking the difference between word frequency at each level of context for targets, we observed that by adding colour to a word we actually decreased the effect of word frequency, which was contrary to our hypothesis.

Given that we hypothesized that the effect of word frequency should have been larger in the coloured condition, such that coloured low frequency words should have had higher claims of recollection, it is important to discuss this further. Perhaps the dual processing theory or the early phase elevated attention hypothesis could account for these results. The dual processing theory suggested that low frequency words were easier to recollect because they were more distinctive than high frequency words. We hypothesized that the addition of a tertiary colour should have created a further advantage for low frequency words, since by doing so, we were making the words even more distinctive. Although our findings were not significant, we did find that for recollection, the mean for LWF/tertiary (M = 5.69) was slightly larger than the mean for LWF/black (M = 5.20). However, the opposite was true for familiarity, whereby the mean for LWF/black (M = 3.70) was larger than the mean for LWF/tertiary (M = 3.50). In overall old responses, the effect of word frequency collapsed over high and low frequency was slightly larger in the black condition (M = 8.90) compared to the tertiary condition (M = 8.85), but according to the dual processing account, the tertiary condition should have had a higher hit rate because it was slightly more distinctive (see Table 3).

Since the dual processing theory was unable to completely account for our results, perhaps the early-phase elevated attention hypothesis would be a better candidate. According to the early-phase elevated attention hypothesis, low frequency words were harder to process, which meant more attention was devoted to them, and this additional attention produced the memory advantage seen in recognition tests (Malmberg and Nelson, 2003; Criss and Malmberg, 2008). Criss and Malmberg (2008) suggested that both the extra attentional resources during encoding, and the greater the familiarity during retrieval, resulted from the unique features linked to low frequency words. In our experiment, the effect of word frequency collapsed over low and high frequency was larger in the black condition compared to the coloured condition for overall old responses (see Table 3). However, for recollection responses, the coloured condition mean (M = 5.69) was slightly larger than the black condition (M = 5.20) for low word frequency, although this difference was not significant. In other words, by adding tertiary colours to the test items we created a small difference in the distinctiveness of recollected responses that increased the hit rate, which coincided with the early-phase elevated attention hypothesis.

Although colour may have added some distinctiveness, it actually decreased the effects of word frequency, as previously mentioned. Perhaps when the word was black, the participants focused more attention on the word itself because no tertiary colours were present that could take attention away from the word. Given that the false alarm rate was higher for overall LWF/tertiary words (M = 3.50) compared to overall LWF/black words (M = 3.25) perhaps the participants paid more attention to the colour of the words, instead of focusing attention on the word itself, and as a result, memory performance was negatively affected.

One further consideration must be mentioned. That is, the distinctiveness/fluency processing framework (Rajaram & Geraci, 2000). This framework suggested that distinctiveness of processing was important for recollection or remembering, and fluency of processing was essential for knowing or familiarity (Rajaram & Geraci, 2000). Low frequency words contained more contextual details compared to high frequency words, and recollection was influenced by processing the distinctive properties of events (Hunt & McDaniel, 1993). Therefore, familiar responses should be responsive to fluency of processing, and the occurrence of recollection should be based on the processing of distinctive information and not based on fluency of processing (Rajaram & Geraci, 2000).

Interestingly, Jacoby and Dallas (1981) suggested that the absolute fluency of processing was higher for high frequency words compared to low frequency words. Given that familiarity responses were sensitive to fluency of processing, perhaps this was the reason our high frequency words were claimed to be more familiar, compared to low frequency words (See Table 2). Furthermore, assuming the addition of colour affected the fluency of processing, this may explain why the mean hit rate for HWF/black words for familiarity ($M = 4.50$) was higher, since HWF/black words may have higher fluency of processing, compared to HWF/tertiary words ($M = 4.13$), whereby the addition of colour may have impeded the fluency of processing.

Interestingly, the distinctiveness/fluency processing framework posited that conceptual and perceptual salience and distinctiveness would reinforce the occurrence of remembering (Rajaram, 1998). For our experiment, in terms of recollection, the mean hit rate for LWF/tertiary words was larger ($M = 5.69$) compared to the mean hit rate for LWF/black words ($M = 5.20$). Although this interaction was not significant, perhaps the distinctiveness/fluency processing framework could explain this difference in means by suggesting that the coloured low frequency words were more distinctive and perceptually salient, compared to the black low frequency words. Interestingly, if one were to take the difference from the mean hit rate for LWF/tertiary, which was the most distinct and salient condition (M = 5.69) and the mean hit rate for HWF/black, which was the least distinct and salient condition (M = 4.70), there is a difference of almost 1.0 (see Table 1).

Nevertheless, we have yet to explain why there was no significant interaction for context x test x word frequency for recollection responses. Our hypothesis was that low frequency words combined with a tertiary colour would produce a higher hit rate for recollection, compared to high frequency words regardless of colour, or low frequency words that were black. Mulligan, Smith, and Spataro (2015) conducted a study, whereby they manipulated font and colour to assess the attentional boost effect (ABE). The ABE proposed that stimuli together with targets in a detection task would be better remembered than stimuli together with distractors. Although our experiment did not utilize a detection task, the results were intriguing enough to include here.

Consequently, Mulligan et al. (2015) found that context memory for colour and font combined was no better in the target condition than the distractor condition. Furthermore, the target trials did not enhance memory for visual features, in this case font and colour, so there was no evidence of enhanced perceptual processing in tests of context memory. In terms of relating this finding back to our current study, we did not find a large advantage in either recollection or overall old responses for tertiary coloured words collapsed over low and high word frequency (see Table 1 and Table 3). Perhaps this result is not surprising after all, given the findings by Mulligan et al. (2015). Certainly, future research would benefit from investigating the reasoning behind this, since this finding was contrary to our hypothesis.

Overall, our results suggested the following: Memory was above chance for recollection, but was not above chance for familiarity. The implication was that participants were treating their familiarity judgments as confidence judgments, and high frequency words were more familiar, which influenced their familiarity judgments. We found a mirror effect for recollection and for overall old responses (see Table 4 and Table 5). For recollection, low frequency words had a larger hit rate compared to high frequency words, and as a result, when a target word had low frequency, participants were more likely to claim recollection instead of familiarity (see Table 1). Given our results, there was clear evidence that when participants were lacking recollection memory, they were utilizing the familiarity of the words to influence their decision. This illustrated a difference in discrimination between items that were recollected compared to items that were familiar, which supports the Remember-Know paradigm.

We found a trend toward word frequency by test for familiarity, which would have replicated the mirror effect had it been significant. Although this interaction was not significant, participants tended to say a word was familiar if the word was low frequency and it was a target, since the hit rate for low word frequency collapsed over colour and black conditions was larger than the corresponding false alarm rate. However, if the word was high frequency and either a target or a distractor, participants were more likely to say the word was familiar. In terms of local context effects, there was only an interaction between context and word frequency for overall old responses. By taking the difference between word frequency and each level of context for targets we found that by adding colour to a high frequency word we decreased the effects of word frequency, which conflicted with our hypothesis. Our results were not completely explained by either the dual processing theory or the early phase elevated attention hypothesis, but the reason for this may be due to the recent discovery made by Mulligan et al. (2015), which suggested there was no heightened perceptual processing in tests of context memory for colour and font combined.

The novel aspect of our experiment was whether local context and word frequency would affect recognition memory in the Remember-Know procedure. Although we replicated the mirror effect for recollection memory and for overall old responses, we did not find an effect of context. It was possible that we did not find a context effect because our study lacked power, given the fact that we only had 16 participants. Further limitations of our study may include the following: because our participants were tested in a group setting, they sat at varying distances from the screen. Perhaps the overhead projector diluted the tertiary colours when the PowerPoint presentation was shown to the participants, so the tertiary colours were not quite as distinctive as they otherwise could have been. Additionally, there may have been various distractions during the study and test phase that could have affected memory retention and recognition, not to mention we tested the participants in the morning, so they may not have been fully awake which could have affected their memory and attention span.

Future researchers may want to investigate why we were unable to find a significant context effect for recollection and familiarity. Perhaps, as stated previously, the overhead projector diluted the distinctiveness of the tertiary colours we used in the experiment. On the other hand, perhaps the finding by Mulligan et al. (2015) was correct, and context memory for colour and font combined is no better for targets than for distractors. Additionally, future researchers may want to investigate whether colour is adding another dimension to the reading of a word, as evidenced by the stroop effect. The stroop effect occurs when the stimulus impedes with the naming of the letter color. For example, when the word red is written in blue letters, and participants must say the colour of the word, rather than reading what the word says (Burt, 2002). Perhaps the colour of high frequency words is named faster than the colour of low frequency words, and maybe this makes a difference for measures of recollection and familiarity.

Not to mention, perhaps whether or not a word is congruent with its colour affects recollection. For example, in our experiment we used the word "sun" and it was coloured blue-purple. Would it make a difference if the word sun was actually coloured yellow? Maybe there would be a larger context effect if future researchers were able to test the incongruence versus congruence aspect of colours and words by testing one half of the participants with congruent colours and words, and the other half with incongruent colours and words.

In the end, our study replicated a few major findings, such as the mirror effect for measures of recollection and overall old responses. However, our novel hypothesis that colour in combination with low frequency words would produce a further advantage for recollection memory was not supported. If anything, our experiment illustrated why it is so vital to replicate previous studies in the literature. This is especially important, given the fact that we were unable to adequately explain why colour did not provide a significant recollection advantage for low frequency items. When one is unable to replicate previous findings in the literature, it is vastly important to find alternative explanations for the observed results through future research.

References

Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 830-842. doi:http://dx.doi.org/10.1037/0278-7393.28.5.830

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 1-15. doi:http://dx.doi.org/10.1016/S0749-596X(02)00518-1

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

Burt, J. S. (2002). Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1019-1038. doi:http://dx.doi.org/10.1037/0096-1523.28.5.1019

Cook, G.I. Marsh, R.L., & Hicks, J.L. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition, 34,* 240-250.

Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language, 59*(3), 331-345. doi:http://dx.doi.org/10.1016/j.jml.2008.05.002

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*(1), 23–30. doi:10.3758/BF0 3202642

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. Journal of *Memory and Language, 32*, 421—445.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*(3), 306-340. doi:http://dx.doi.org/10.1037/0096-3445.110.3.306

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition, 31*(1), 35-43. doi:http://dx.doi.org/10.3758/BF03196080

Meeks, J. T., Knight, J. B., Brewer, G. A., Cook, G. I., & Marsh, R. L. (2014). Investigating the subjective reports of rejection processes in the word frequency mirror effect. *Consciousness and Cognition: An International Journal, 24*, 57-69. doi:http://dx.doi.org/10.1016/j.concog.2013.12.007

Mulligan, N. W., Smith, S. A., & Spataro, P. (2015). The attentional boost effect and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* doi:http://dx.doi.org/10.1037/xlm0000183

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

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5*(1), 71-78.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 1070-1074. doi:http://dx.doi.org/10.1037/0278-7393.26.4.1070

Steyvers, M., & Malmberg, K.J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 760-766.

Table 1

Number of Hits and False Alarms for Recollection

	LWF	HWF	Mean
		Old Items (Hits)	
Tertiary	5.69	4.38	5.04
Black	5.20	4.70	4.95
Mean	5.45	4.54	5.00
		New Items (False Alarms)	
Tertiary	0.44	1.70	1.07
Black	0.50	2.20	1.35
Mean	0.47	1.95	1.21

Table 2

Number of Hits and False Alarms for Familiarity

	LWF	HWF	Mean
		Old Items (Hits)	
Tertiary	3.50	4.13	3.82
Black	3.70	4.50	4.10
Mean	3.60	4.32	4.00
		New Items (False Alarms)	
Tertiary	3.06	4.44	3.75
Black	2.75	4.56	3.66
Mean	3.00	4.50	3.71

Table 3

Overall Number of Hits and False Alarms Collapsed over Recollection and Familiarity

	LWF	HWF	Mean
		Old Items (Hits)	
Tertiary	9.20	8.50	8.85
Black	8.90	8.90	8.90
Mean	9.05	8.70	8.87
		New Items (False Alarms)	
Tertiary	3.50	6.13	4.82
Black	3.25	6.75	5.00
Mean	3.38	6.44	4.91

Table 4

Mirror Effect of Word Frequency for Hits and False Alarms for Recollection

	Old Items (Hits)		New Items (False Alarms)	
Word Frequency	Low	High	Low	High
	5.45	4.53	0.47	1.94

Table 5

Mirror Effect of Word Frequency for Overall Number of Hits and False Alarms Collapsed over Recollection and Familiarity

	Old Items (Hits)		New Items (False Alarms)	
Word Frequency	Low	High	Low	High
	9.03	8.69	3.38	6.44