2016

The Effect of Mixed Font Items on Lexical Decision Performance

Arielle B. Cenin

THE EFFECT OF MIXED FONT ITEMS ON LEXICAL DECISION PERFORMANCE

ARIELLE B. CENIN

Bachelor of Arts in Psychology

Ursuline College

January 2014

submitted in partial fulfillment of requirements for the degree

MASTER OF ARTS IN PSYCHOLOGY

at the

CLEVELAND STATE UNIVERSITY

May 2016

We hereby approve this thesis for

Arielle B. Cenin

Candidate for the Master of Arts in Psychology degree for the

Department of Psychology

and the CLEVELAND STATE UNIVERSITY

College of Graduate Studies

Thesis Chairperson, Albert F. Smith

Department & Date

Methodologist, Conor T. McLennan

Department & Date

Eric S. Allard

Department & Date

Kenneth E. Vail

Department & Date

Student's Date of Defense: 05/11/16

Acknowledgments

Thank you to Dr. Albert Smith for his support as my advisor. Also, thank you to Maryam Assar, Kristyn Oravec, and Hannah Princic for assistance with data collection and experiment programming.

THE EFFECT OF MIXED FONT ITEMS ON LEXICAL DECISION PERFORMANCE

ARIELLE B. CENIN

ABSTRACT

The multistream model of word perception (Allen, Smith, Lien, Kaut, & Canfield, 2009) suggests that word identification generally involves whole-word information, but that when the orthographic form of a letter string is not standard, processing occurs analytically and is slower. For example, within-item case transitions slow responses in lexical decision experiments, in which participants are required to decide if a letter string is or is not a word; a within-item font transition may have a similar effect. Letters within a font are distinct yet related, and are constrained on several parameters to facilitate processing (Sanocki & Dyson, 2012). Font tuning allows design commonalties to be utilized by the perceptual system when processing subsequent items, and changes in font slow processing because the translation rules cannot be carried over (Walker, 2008). We conducted two experiments to investigate the effect of font variation on lexical decision performance. Experiment 1 addressed whether between-item font variation interferes with judgments of lexicality. Planned contrasts showed a marginal difference in response times between pure-font and intermixed-font blocks ($t(1, 23)= 1.45$, $p= 0.07$). Although the results do not pose a strong challenge to the idea that decisions on lexicality are not interfered with by random trial-to-trial variation in font, response times in intermixed font blocks tended to be slower than responses in pure font blocks. Experiment 2 investigated the effect of within-item font transition on lexical decision performance. The significant main effect of font

homogeneity ($t(1, 23)= 1.76$, $p= 0.04$) showed that responses to heterogeneous font items were slower than responses to homogeneous font items. The results of Experiment 2 supported the hypothesis that a within-item font transition slows lexical decision performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The process of visual word recognition has undergone much debate. Most theorists favor the analytic approach, and argue that forming representations of letters is necessary for forming representations of words (e.g., McClelland & Rumelhart, 1981; Besner & Johnston, 1989; Dehaene, Cohen, Sigman, & Vinckier, 2005; Grainger & Jacobs, 1996). Whereas analytic models reason that whole-word information is not necessary for processing, holistically biased models propose that multiple channels are responsible for processing both whole-word (holistic) and letter-level (analytic) information (Allen, Smith, Lien, Kaut, & Canfield, 2009). These channels, each sensitive to certain spatial frequency information (e.g., high, low) process in parallel. Thus, the speed at which an item is identified depends on the familiarity of its orthographic form. For example, compared to homogeneous case items (e.g., travel), items with a single case transition (e.g., TRAvel) have been shown to impair lexical decision performance (Allen, Smith, Lien, & Watt, in preparation). It is unclear how a straightforward analytic model would account for observed processing differences between mixed-case and homongeneous-case stimuli, whereas such differences are consistent with predictions of holistically biased models.

If a case transition impairs lexical decision performance because the case transition disrupts normal orthographic form, then a within-item font transition may have the same effect. Font designs vary on several parameters, thus making different fonts appear dissimilar in weight, contrast, size, and angle (Sanocki & Dyson, 2012). As such, the primary question to be addressed in this thesis is what impact a within-item font transition has on lexical decision performance. First, models of visual word recognition are reviewed. Next, font design parameters will be discussed. Design parameters are followed by a review of font tuning. Then, literature on mixing font will be reviewed. The Introduction will culminate in the description of the experiments.

Models of Visual Word Recognition

An important issue in visual word recognition is whether words are formed on the basis of abstract letter units or whether they are formed on the basis of global word shape. Although early research showed that words might be identified by the use of word shape (e.g., Cattell. 1886), most theorists currently favor models in which words are formed analytically from component letters (Forster, 1976; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Perea & Rosa, 2002). In general, analytical models suggest that information about the physical stimulus is lost early in the process of word recognition, so the particular form a letter takes is irrelevant to processing. Therefore, analytical models tend to predict that performance on items with case transitions should not be slower than the slower of homogeneous uppercase and lowercase. Additionally, the effect, if any, of mixed-case items should be the same for words and nonwords. However, others argue that word shape plays a role in visual word recognition and words may be identified, under certain conditions, on the basis of holistic properties (Allen, Wallace, & Weber, 1995; Healy & Cunningham, 1992; Perea & Rosa, 2002).

Allen, Smith, Lien, Kaut and Canfield (2009) proposed the multistream model, according to which words can be formed either via word-level codes, or via letter-level codes as in analytical models. According to the multistream model, the spatial frequency information of a word is the basic unit of analysis. As such, words that are orthographically familiar (e.g., high frequency words) can be identified by the fast, low frequency sensitive, word-level channel, whereas words that are not orthographically standard (e.g., words that contain case transitions) must be processed by the slow, high frequency sensitive, letter-level channels. As such, the time required to identify a word may provide evidence about the mechanisms involved. For example, in lexical decision experiments in which participants decide whether each presented letter string is or is not a word, responses to homogeneous case strings (e.g., travel, TRAVEL) are faster than responses to mixed-case items (e.g., traVEL and TRAvel). From the perspective of the multistream model, reaction time is slower on mixed-case items because the orthographic form is not standard, thus the slower, letter-level streams win the race to process the high spatial frequency information contained in the letter string (Allen et al., 2009).

The slower reaction times for mixed-case than homogeneous-case words may also be due to differences in attentional priority. For example, research suggests that slowed processing of mixed-case items is caused by the size difference between the uppercase and lowercase letters. With mixed-case words, differences in letter size may disrupt the formation of word units and encourage inappropriate perceptual grouping of same-size or same-case letters. That is, uppercase letters may disrupt whole-word processing because they appear larger than the lowercase letters in a word. For example, Humphreys, Mayall, and Cooper (2003) developed the "buried-word task" (i.e., PIG in sPrInG) in order to investigate the influence of case transition on attention to subparts of words. The results showed that case mixing impaired the identification of whole-words, but facilitated the identification of buried words, relative to when buried words were presented in homogeneous-case letter strings.

Font Design

The individual letters within a high-quality font type are designed to be distinct yet related in order to facilitate reading (Sanocki & Dyson, 2012). Distinctiveness allows each letter to be easily discriminated from others in the alphabet. However, relatedness is equally important in font design as balance and uniformity across letters is necessary for whole-word processing (Carter, Day, & Meggs, 1985; Cheng, 2005; Sanocki & Dyson, 2012). Thus, the letters within a font type are constrained on several parameters in order to achieve cohesiveness; this is done so that commonalties may be utilized by the perceptual system when processing subsequent letters. This mechanism is termed font tuning and has been shown to increase processing efficiency (Walker, 2008). For example, letters within a font share similar features such as letter weight, contrast between strokes, angle of stress, and the reference frame (Sanocki & Dyson, 2012; see Figures 1 and 2). Letter weight depends on the overall thickness of the strokes in a letter, and can be either heavy or light. Contrast between strokes refers to the difference between the thinnest and thickest strokes in a letter. Angle of stress is shown in the contrast between round strokes of the letter, and can be vertical or oblique. The reference frame determines the overall size of a font, and governs the cap height, x-height, and baseline of letters. These design parameters often vary among fonts. As a result, when different fonts in the same point size are mixed, the letters appear diverse in terms of reference frame, weight, contrast, and angle.

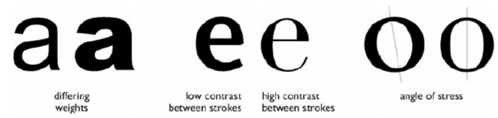

Figure 1. Examples of letter weight, contrast between strokes, and angle of stress. (From Sanocki and Dyson, 2012).

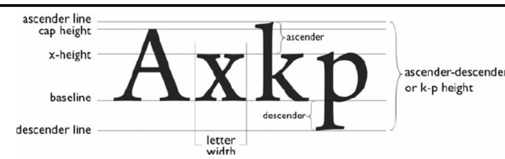

Figure 2. Example of the systematic reference frame. (From Sanocki and Dyson, 2012).

Although mixing font within a letter string increases the distinctiveness of individual letters, it slows the identification of the letter string (Gauthier, Wong, Hayward, & Cheung, 2006). Therefore, mixing font within a letter string is likely to produce the same pattern of results as shown with mixed-case items. In the experiments conducted for this thesis, we investigated the effect of within-item font changes on lexical decision performance utilizing two dissimilar fonts. The fonts chosen for these experiments (i.e., Garamond & **Arial Black**) are dissimilar in numerous characteristics. Thus, a within-item font transition using these fonts may make font tuning more difficult, as well as disrupt holistic processing. That is, mixed-font items may require letter-level processing and slow lexical decision performance.

Font Tuning

The study of font tuning is the study of how changes in font are processed. For example, when an individual begins to read text presented in a novel font type, he or she must determine the "translation rules" for that particular font. This is likely to slow processing and lead to more errors (Walker, 2008). Consequently, when font is mixed between strings, the degree of similarity between fonts is an important factor (Sanocki, 1992; Walker, 2008; Sanocki & Dyson, 2012). For example, when font types with similar design parameters are mixed between items, the reader is able to use similar translation rules to continue fluent reading. However, when the font types are more dissimilar, the reader may struggle with letter identification and processing may be slowed.

Walker (2008) used two dissimilar fonts, **Cooper Black** and *Palatino Italic*, in a variant of a lexical decision task in which each target word or nonword string was presented with a distractor consonant string; the two strings were arranged vertically, and which stimulus was in the upper position was unpredictable. On any trial, the target string and the irrelevant consonant string were either in the same font or different fonts. The function of the consonant string was to determine if "the translation rules applied to one portion of text are removed from working memory before the next portion of text is dealt with" (Walker, 2008, p. 1034). That is, by presenting the target string and distractor consonant string simultaneously, the time between presentations was reduced; this allowed for pairs of target strings and consonant strings to appear with equal probability in either of the two fonts, and for successive pairs to be equally likely to appear in the same font or in different fonts. Participants were instructed to avoid responding to the consonant string, and to decide if the target stimulus was or was not a word. Walker found that lexical decision responses were quicker when the target and distractor strings were presented in the same font than in different fonts, from which he argued that slower responses to targets, when the target and distractor were in different fonts, was because the differences in font design made font tuning more difficult for the reader.

Tran, Joyce, and Cottrell (2004) performed multidimensional scaling on dissimilarities among 15 fonts; the dissimilarity measure was based on a principal components analysis of filtered images of all letters in each of the fonts. Their two-dimensional solution is shown in Figure 3. Distances between fonts “were defined as one minus the average correlations between their corresponding letters” (Tran et al., 2004, p. 3). The results showed that fonts such as XeroxSerifNarrow and Mirror are dissimilar in design. On one end of the spectrum, font types have characteristics such as heavy letter weight, low contrast between strokes, vertical angle of stress, and a large reference frame. On the contrasting end of the spectrum, fonts have low letter weight, high contrast between strokes, oblique angle of stress, and a smaller reference frame. Sanocki & Dyson (2012) posited that when fonts with highly similar design characteristics are

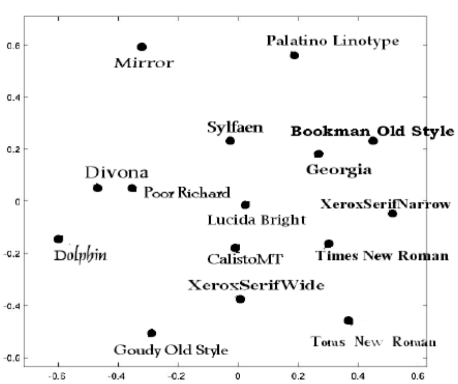

Figure 3. Multidimensional scaling solution of similarities among 15 fonts. (From Tran, Joyce, and Cottrell, 2004).

presented together, font tuning is facilitated. For example, fonts such as Optima and Bookman Old Style have similar global information. Thus, when a within-item transition occurs with these two fonts, the translation rules from one can be applied to the other, which should allow a reader to rapidly identify the item. However, when the fonts are greatly dissimilar, font tuning would be more difficult and word identification may be slowed. Additionally, the extent to which processing is slowed may depend on the degree of dissimilarity between two fonts (Sanocki & Dyson, 2012).

Moret-Tatay and Perea (2011) conducted lexical decision experiments in which stimulus items were in Lucida Bright and Lucida Sans— two fonts from the same family. The fonts differed only in the absence or presence of serifs. Words and nonwords were presented in either pure-font or mixed-font blocks. In mixed-font blocks, both fonts were presented in random order. The results showed no difference in reaction time on mixed-font blocks compared to pure-font blocks. However, in the pure-font blocks, reaction times for words in sans serif font were slightly faster than those for words in serif font This result slightly contradicts the theory of font tuning and provides evidence that between-item variation in font does not necessarily interfere with lexical decision performance. The purpose of our Experiment 1 is to ascertain whether this is the case for the fonts that we have chosen (i.e., Garamond and **Arial Black**) to investigate the effects of within-item font transitions on lexical decision performance.

Font Mixing

Gauthier, Wong, Hayward, & Cheung (2006) asked participants to search for a target letter within 10 x 10 matrices. Participants began the search task by identifying the first letter in the matrix as the target letter. Next, they were asked to scan the matrix left to right, top to bottom until they found the target letter. The letter immediately following was the new target letter. This target scanning process was continued until the end of the matrix was reached. In their experiment, five different fonts (i.e, Georgia, Croissant, Larabi, Trebuchet, and Angelus) were used in three different types of matrices: baseline, regular, and mixed. In baseline matrices, all 100 letters were in the same font, with matrices in each of the five fonts; in regular matrices, the letters in each row were in the same font, but font changed between rows; and in mixed matrices, the font assigned to each position was random. The results showed that searches were faster and more accurate when all letters in the matrix were in the same font than when the letters in the matrix varied in font. Furthermore, the search was significantly faster and more accurate in the regular matrix condition than in the mixed matrix condition.

Sanocki (1987) developed two fonts that were similar in overall letter size, but differed on several properties including the extent and details of terminations, basic shape of parts, line thickness, and spatial dimensions. Participants completed a letter-nonletter task in which they were asked to discriminate between strings with all letters and strings with one nonletter. (Nonletters were made by either adding or removing one segment of a letter.) Font was either consistent within strings or mixed within strings. Results showed faster response times for same-font strings than for mixed-font strings. Furthermore, Sanocki (1992) asked participants to complete a letter-nonletter task in which strings were presented in two fonts that differed in shape, size, and presence or absence of serifs. Results showed that reaction time was slowed significantly when transitioning from gothic (i.e., sans serif) to serif font. It was suggested that serifs added a higher level of complexity, which resulted in slower processing compared to items in gothic font.

Present Experiments

The purpose of these experiments was to determine if a within-item font transition interferes with lexical decision performance. This question would be most interestingly addressed under conditions in which between-item variation in font does not impair performance; the purpose of Experiment 1 was to investigate the effect of between-item variation in font. Moret-Tatay and Perea (2011) found that between-item variations in font did not interfere with lexical decision performance. It is of interest to see if similar results are found with the fonts that we have chosen, given that these fonts are not part of the same font family.

The fonts we have chosen for the present experiments, Garamond and **Arial Black**, are dissimilar on several parameters, including the absence or presence of serifs, letter weight, contrast between strokes, and the reference frame. When presented in the same point-size, Arial Black appears significantly larger than Garamond. However, the proportional discrepancy between these two fonts has been addressed in the current experiment so that items in Garamond are the same size as those in Arial Black (i.e., items in Garamond are set to 36-point font while items in Arial Black are set to 28-point font). This is a novel approach, as most research on mixing high quality font has not attempted to manipulate point size to make spatial parameters appear equal (e.g. Walker, 2008; Gauthier et al., 2006). Previous research has shown that differences in spatial parameters generally slow processing and word identification in mixed-font items (Sanocki & Dyson, 2012). By making contrast between strokes, letter weight, and the absence or presence of serifs the primary discrepancies, we hope to gain further insight on how high spatial frequency information is processed in mixed-font items.

The purpose of Experiment 1 was to determine whether variation in font across trials interferes with judgments of lexicality. Four blocks of stimuli were presented. In two of the blocks, all stimuli were the same font—Arial Black in one block and Garamond in the other—and in the other two blocks, stimuli of these two fonts were randomly intermixed. The purpose of Experiment 2 was to investigate the effect of within-item font transition on lexical decision performance. Each block consisted of randomized trials presented in both homogeneous-font and mixed-font format.

CHAPTER II

EXPERIMENT 1

Method

Participants. Undergraduate Psychology students at Cleveland State University participated in this experiment. They were recruited through SONA Research Participation System, and were compensated with research credit upon completion of the experiment. Participants were asked to report whether English is their first language.

Materials. A master list of 256 words was used; from each word, a corresponding pseudoword was created by changing one letter (e.g., “down” and “dowy”). Words were 4- and 6-letters long, and were either high or low frequency. The Kučera and Francis (1967) frequencies per million are as follows: high frequency, 4- letter: *M*= 430, *SD*= 243; high frequency, 6- letter: *M*= 186, *SD*= 145; low frequency, 4- letter: *M*= 19, *SD*= 5; low frequency, 6- letter: *M*= 20, *SD*= 6. See Appendix A for the complete stimulus list.

For each item on the master list, two image files were created: one in Arial Black 28 point font (**AA**) and one in Garamond 36 point font (GG). For each participant, 128 words were randomly selected from the master list; these and the complementary 128 pseudowords were used. Thus, no participant saw both a word and its corresponding pseudoword. Unique stimulus lists were generated for each participant containing a random half of the items from each of the eight stimulus classes formed by crossing two levels of Lexicality (word, pseudoword), two levels of Frequency (high, low), and two levels of Length (4-, 6- letter). These were assigned at random to conditions and lists, subject to the constraint that each pure font list had eight items from each of the eight Lexicality x Frequency x Length classes, and each of the intermixed list had four items from each of the 16 Font x Lexicality x Frequency x Length classes. Figure 4 illustrates examples of parts of the stimulus lists for Experiment 1.

terror	**casual**	tennis
repeir	**slanet**	**member**
balt	**office**	**snap**
little	**achoss**	frield
dowy	**bame**	**told**
bare	**rent**	fark
back	**cety**	**slanet**
areund	**used**	**cond**

Figure 4. Examples of pure-font and intermixed stimulus lists for Experiment 1. The left column shows part of a GG (pure Garamond) list; the center column shows part of an AA (pure Arial Black) list; and the third column shows part of an intermixed list.

Design. For Experiment 1, a 2 (Lexicality: word, pseudoword) x 2 (Block Type: pure-font, intermixed-font) x 2 (Font: Arial Black, Garamond) x 2 (Length: 4-, 6- letter) x 2 (Frequency: high, low) within-subjects design was used. The variables measured were response time (in ms) and accuracy. For each participant, there were two pure-font blocks—one Arial Black and one Garamond—and two intermixed font blocks in which the font of items varied randomly between trials, subject to the constraints imposed by the construction of trial blocks (described above). For each participant, the order of pure and intermixed blocks alternated, and this was counterbalanced across participants: For half of the participants, the first trial block was a pure-font block; for the other half, the first trial block was an intermixed-font block. Within this balancing plan, for half of the participants, the Arial Black block was the first pure-font block; for the other half of the participants, the Garamond block was the first pure-font block.

*Procedure***.** Participants performed the lexical decision task on a desktop computer. SuperLab 4.0 was used to run the experiment and collect data for analysis. The participant was instructed to, as quickly and accurately as possible, decide if the letter string presented on each trial was a word or not a word. The participant was asked to press the left (red) button if the string was not a word, and the right (green) button if the string was a word. Each trial began with a 300 ms warning cross, followed by a 300 ms blank, followed by a stimulus that was presented until the participant responded; response time was measured from the onset of the stimulus to the response. Following a response, there was a 500 ms blank.

Each participant first completed a series of practice blocks. The practice consisted of four blocks—one in Garamond, one in Arial Black, and two with intermixed font— with eight trials each, where the order of pure and intermixed blocks alternated. For half the participants, the first block was pure font, and for the other half the first block was intermixed; this was done to match the construction of the blocks in Experiment 1. Participants were required to score 87% correct or above on the practice in order to continue on to Experiment 1. If the participant did not reach this criterion, he or she was asked to complete a second practice. If the participant did not reach the cutoff for accuracy after the second practice, their data was not included in the analysis. After the participant completed the four experimental blocks, he or she was debriefed.

Results

Data were collected from 28 participants in order to obtain analyzable data from 24. One participant did not satisfy the pre-established criterion in the practice blocks (i.e., 87%) and three participants' data were not included in the analysis due to programming errors.

For each participant, for each of words and pseudowords, error rate and median response times on correct trials were found for each of the 16 conditions that resulted from crossing completely two levels of block type, two levels of font, two levels of frequency, and two levels of length. (For pseudowords, frequency was defined by the frequency of the word from which the pseudoword was created.)

Response times to words. Table B1 in Appendix B shows means and standard deviations of response times for words for each Block Type x Frequency x Length x Font conditions. Fifteen contrasts that correspond to the four main effects and all interactions of an analysis of variance were calculated (See Appendix C for contrast tables). However, we were interested in and planned to examine only the four main effects—Block Type (pure vs. intermixed), Font (Arial Black vs. Garamond), Length (4- vs. 6- letter), and Frequency (high vs. low)—and the two-way interaction of Block Type x Font. For contrasts that tested directional questions, t statistics and one-tailed p-values are reported; for contrasts that tested non-directional questions, F statistics are reported. The main effects of frequency and length were analyzed to see if effects common in results of previous lexical decision experiments are also in these data. Specifically, we were interested in whether responses to 6-letter items were slower than those to 4-letter items, and whether responses to low frequency words were slower than those to high frequency words. The main effect of Length ($t(23)= 3.86$, one sided, $p= 0.0004$; $\eta_p^2 = 0.39$) was significant, with slower response times to 6-letter words ($M= 753$, $SD= 211$) than to 4-letter words ($M= 696$, $SD= 124$) (see Figure 5). The main effect of Frequency ($t(23)= 6.28$, one sided, $p=.000001$; $\eta_p^2 = 0.63$) was also significant, with slower responses to low frequency words ($M= 756$, $SD= 208$) than to high frequency words ($M= 684$, $SD= 123$) (see Figure 6). These results were consistent with those of many other lexical decision experiments.

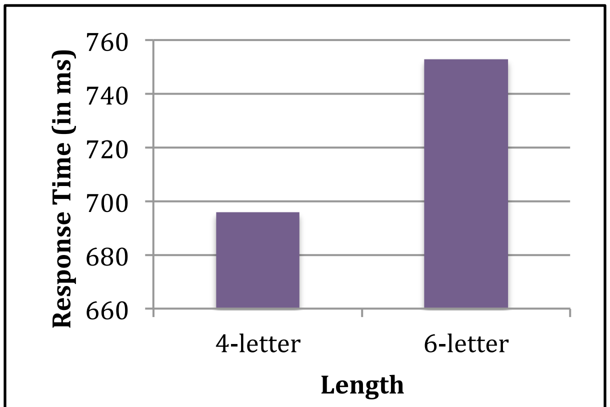

Figure 5. Mean response times (in ms) for words in Experiment 1, by length.

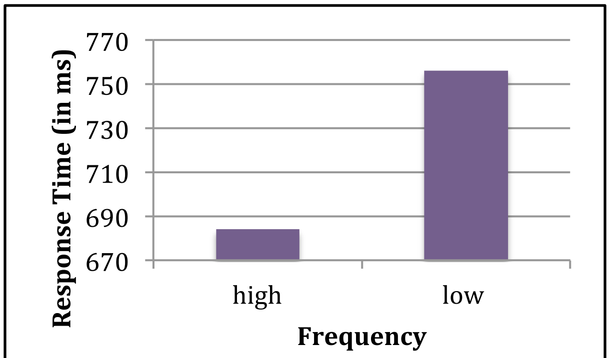

Figure 6. Mean response times (in ms) for words in Experiment 1, by frequency

For Experiment 1, the main question was whether responses in the intermixed-font blocks are slower than responses in pure-font blocks. Although the effect of Block was not statistically significant ($t(23)= 1.46$, one sided, $p= 0.08$; $\eta_p^2 = 0.08$), responses in the intermixed-font blocks tended to be slower ($M = 735$, $SD = 191$) than those in the pure-font blocks ($M = 714$, $SD = 158$). Font ($F(1, 23)= 0.73$, $p=$

0.40; η_p^2 = 0.03), and the Block Type x Font interaction (F(1, 23)= 0.24, p= 0.63; η_p^2 = 0.01) were not statistically significant. The remaining 10 contrasts were combined into a residual term, which was not statistically significant (F(10, 230)= 1.20, p= 0.29; η_p^2 = 0.05).

Error rates to words. Table B2 in Appendix B shows means and standard deviations of error rates for words for each Block Type x Frequency x Length x Font condition. For error rates for words, the main effect of Frequency (t(23)= 4.01, one sided, p= .0003; η_p^2 = 0.41) was significant, in which error rate was lower for high frequency than low frequency words (high frequency: M= 0.02, SD= 0.05; low frequency: M= 0.04, SD= 0.08) (see Figure 7). The effects of Block Type (t(23)= 0.54, one sided, p= 0.30; η_p^2 = 0.01), Font (F(1, 23)= 2.62, p= 0.12; η_p^2 = 0.10), Block Type x Font (F(1, 23)= 2.33, p= 0.14; η_p^2 = 0.09), and Length (t(23)= 1.51, one sided, p= 0.07; η_p^2 = 0.09) were not statistically significant (4- letter: M= 0.03, SD= 0.07; 6-letter: M= 0.04, SD= 0.07) . However, the residual term (F(10, 230)= 2.34, p= 0.01; η_p^2 = 0.09) was statistically significant. Therefore, the residual was decomposed into the 10 1 degrees of freedom components representing all two-way (except Block Type x Font) and three-way interactions of Block Type, Font, Frequency, and Length, and the four-way interaction of these variables. The following have p-values of 0.05 and

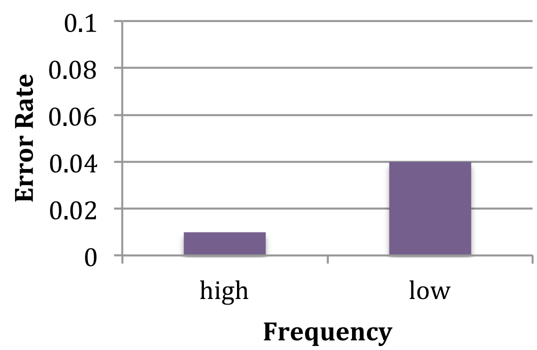

Figure 7. Mean error rates for words in Experiment 1, by frequency.

below: Frequency x Length ($F(1, 23)$= 4.10, p= 0.04; η_p^2 = 0.15), Block Type x Frequency x Length ($F(1, 23)$= 6.28, p= 0.02; η_p^2 = 0.21), and Block Type x Font x Frequency x Length ($F(1, 23)$= 4.33, p= 0.05; η_p^2 = 0.16).

Response times to pseudowords. See Table B3 in Appendix B for means and standard deviations for response times to pseudowords for each Block Type x Frequency x Length x Font condition. For response times to pseudowords, the main effect of Length was significant ($t(23)$= 3.23, one sided, p= 0.002; η_p^2 = 0.31), with slower responses to 6- letter pseudowords than to 4- letter pseudowords (6- letter: M= 1052, SD= 656; 4- letter: M= 942, SD= 448) (see Figure 8). The effects of Block Type ($t(23)$= 0.70, one sided, p= 0.25; η_p^2 = 0.02), Font ($F(1, 23)$= 0.00003, p= 0.10; η_p^2 = 0.000001), Block Type x Font ($F(1, 23)$= 0.11, p= 0.74; η_p^2 = 0.005), and Frequency ($F(1 ,23)$= 2.48, p= 0.13; η_p^2 = 0.10) were not statistically significant. The residual term also was not statistically significant ($F(10, 230)$= 0.60, p= 0.81; η_p^2 = 0.03).

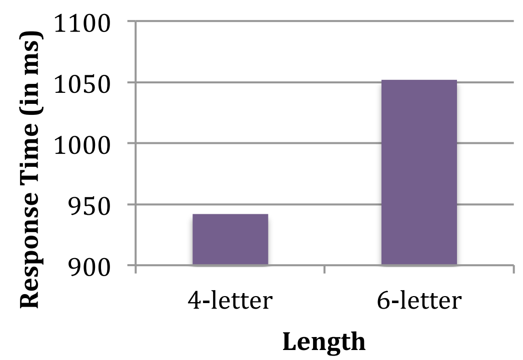

Figure 8. Mean response times (in ms) for pseudowords in Experiment 1, by length.

Error rates to pseudowords. See Table B4 in Appendix B for means and standard deviations for error rates to pseudowords for each Block Type x Frequency x Length x Font condition. For error rates to pseudowords, the main effects of Block Type ($t(23)$= 1.74, one sided, p= 0.05; η_p^2 = 0.12) and Frequency were significant ($t(23)$= 1.68, one sided, p= 0.05; η_p^2 = 0.11). For Block Type, error rates to pseudowords were greater in pure font blocks than in intermixed font blocks (pure: M= 0.09, SD= 0.12; intermixed: M= 0.07, SD= 0.11) (see Figure 9). For Frequency, error rates to pseudowords were higher for "low frequency" (M= 0.08, SD= 0.12) than for "high frequency" items (M= 0.07, SD= 0.12) (see Figure 10). The effects of Font ($F(1, 23)$= 0.83, p= 0.37; η_p^2 = 0.03), Block Type x Font ($F(1, 23)$= 0.25, p= 0.62; η_p^2 = 0.01), and Length ($t(23)$= 0.71, one sided, p= 0.24; η_p^2 = 0.02) were not statistically significant. Also, the residual term was not significant ($F(10, 230)$= 1.47, p= 0.15; η_p^2 = 0.06).

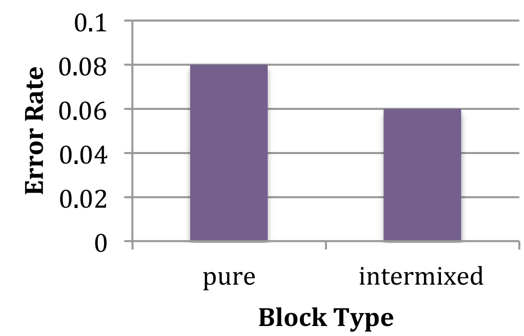

Figure 9. Mean error rates to pseudowords in Experiment 1, by block type.

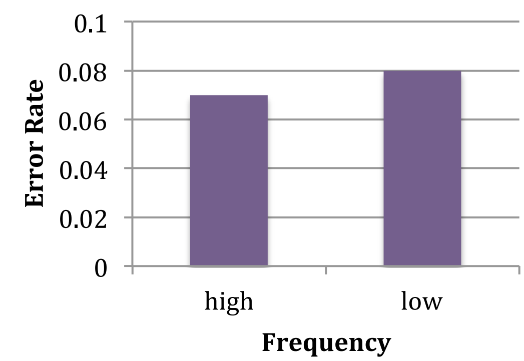

Figure 10. Mean error rates for pseudowords in Experiment 1, by frequency.

Discussion

The significant main effects of Length and Frequency were consistent with those of many other lexical decision experiments. That is, responses to 6- letter words were slower than those to 4- letter words, and responses to low frequency words were slower than those to high frequency words. Furthermore, the main effect of Block Type was not statistically significant, nor was the interaction between

Block Type x Font. Although on average, there was a tendency for responses to be faster in pure than in intermixed blocks, there was also tendency for responses to be more errorful in pure than intermixed blocks. Therefore, the results of Experiment 1 showed that random trial-to-trial variation in font did not, overall, affect performance, and that there were no overall differences between the fonts. These results are consistent with the idea that, for the fonts studied, decisions about lexicality are not substantially affected by random trial-to-trial variation in font.

CHAPTER III

EXPERIMENT 2

Method

*Participants***.** Undergraduate Psychology students at Cleveland State University participated in this experiment. They were recruited through SONA Research Participation System, and were compensated with research credit upon completion of the experiment. Participants were asked to report whether English is their first language.

Materials. The same master list of 256 words and corresponding pseudowords used for Experiment 1 was used. (See Appendix A for the complete stimulus list). Items were assigned to lists subject to the constraint that each list contained two items from each of the Lexicality x Font x Frequency x Length classes. There were four blocks in which all item types were randomly intermixed.

For each item on the master list, four image files were created: one in Arial Black 28 point font (**AA**); one in Garamond 36 point font (GG); one transitioning from Arial Black to Garamond (**A**G); and one transitioning from Garamond to Arial Black (G**A**).

Unique stimulus lists were generated for each participant: For each participant, 128 words were randomly selected from the master list, with 32 from each of the four Frequency (high, low) x Length (4-, 6- letter) classes; these and the complementary 128 pseudowords were used. Thus, no participant saw both a word and its corresponding pseudoword. From each class items were allocated at random to the four levels of Font (Arial Black, Garamond, Arial Black-to-Garamond, Garamond-to-Arial Black); we call pure Garamond and pure Arial Black items homogeneous-font stimuli, and stimuli that transitioned from Garamond-to-Arial Black or from Arial Black-to-Garamond heterogeneous-font stimuli. Figure 11 illustrates examples of parts of the stimulus list for Experiment 2.

does	dake	**pott**
idea	ur**gy**	absorb
di**sk**	roar	**candle**
told	**ou**ce	lo**ny**
ouce	**cond**	**vic**tim
wake	**mom**ert	do**or**
dantal	mem**ber**	**fo**me
fethod	claced	**mist**

Figure 11. Example of stimulus list for Experiment 2. Homogeneous font items and heterogeneous font items were randomly intermixed.

Design. A 2 (Lexicality: word, pseudoword) x 4 (Font: **AA**, GG, **A**G, G**A**) x 2 (Frequency: high, low) x 2 (Length: 4-, 6- letter) within-subjects design was used. The variables measured were response time (in ms) and accuracy.

*Procedure***.** The apparatus and procedure were identical to those in Experiment 1.

Results

Data were collected from 29 participants in order to obtain analyzable data from 24. Four participants did not satisfy the pre-established criterion in the practice blocks (i.e., 87%) and one participant's data were not included in the analysis due to a programming error.

Experiment 2 addressed whether a within-item font transition slows lexical decision performance. Response times and error rates were examined separately for words and pseudowords. A set of 15 contrasts was performed to examine the effect of font homogeneity (homogeneous vs. heterogeneous), and then within each of these stimulus types, the effects of font, frequency, length, and the 2- and 3-way interactions of these three variables. We were specifically interested in the main effect of Stimulus Type (homogeneous vs. mixed): Are responses to heterogeneous font stimuli (e.g., do**wn**, **do**wn) slower than those to homogeneous font stimuli (e.g., down, **down**)? Then, separately for homogeneous and heterogeneous font trials, we tested the effects of the two font instantiations (AA vs. GG, AG vs. GA), frequency (high vs. low), and length (4- vs. 6- letter). (See Table C2 of Appendix C for the contrasts). As planned, we aggregated the 2- and 3-way interactions of font, frequency, and length at each level of font homogeneity into a residual that we also tested.

*Response times to words***.** See Table B5 in Appendix B for means and standard deviations for response times to words for each Font x Frequency x Length condition. The contrast that compared homogeneous and heterogeneous font items was statistically significant ($t(23)= 1.77$, one sided, $p= 0.045$; $\eta_p^2 = 0.12$), which showed that a within-item font transition affected performance. This finding supported the hypothesis that a within-item font transition slows lexical decision performance, with slower responses to heterogeneous font items than to homogeneous font items (homogeneous: *M*= 672, *SD*= 108; heterogeneous: *M*= 678, *SD*= 142) (see Figure 12). There was a significant effect of frequency within each of homogeneous fonts and heterogeneous fonts (Frequency within homogeneous fonts: $t(23)= 2.43$, one sided, $p= 0.01$; $\eta_p^2 = 0.20$; Frequency within heterogeneous fonts: $t(23)= 2.91$, one sided, $p= 0.004$; $\eta_p^2 =$ 0.27) (see Figure 13). Responses to low frequency words were slower than responses to high frequency words at each level of font homogeneity (for homogeneous, for high frequency: *M*= 654, *SD*= 102; low frequency: *M*= 682, *SD*= 113; for heterogeneous, for high frequency: *M*= 653, *SD*= 122; low frequency: *M*= 694, *SD*= 155). Finally, the contrast for word length within heterogeneous fonts ($t(23)= 5.10$, one sided, $p= .00002$; $\eta_p^2 = 0.53$) was significant, but not within homogeneous fonts ($t(1, 23)= 1.22$, one sided, $p= 0.12$; $\eta_p^2 = 0.06$); responses to 6- letter words were slower than responses to 4- letter words (for heterogeneous, for 6- letter: *M*= 689, *SD*= 159; 4- letter words: *M*= 657, *SD*= 119; for homogeneous, for 6- letter: *M*= 672,

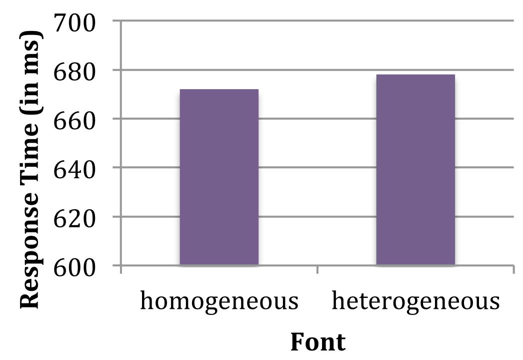

Figure 12. Mean response times (in ms) to words in Experiment 2, by font.

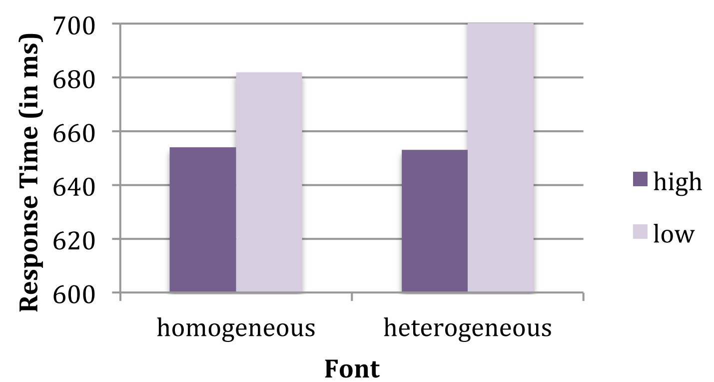

Figure 13. Mean response times to words (in ms) in Experiment 2, by frequency.

SD= 115; 4- letter: *M*= 663, *SD*= 101) (see Figure 14). However, the contrasts comparing font instantiation within homogeneity and heterogeneity were not significant (AA vs. GG: (F(1, 23)= 0.11, p= 0.74; η_p^2 = 0.005; AG vs. GA: (F(1, 23)= 0.09, p= 0.76; η_p^2 = 0.004). The remaining eight contrasts were combined into a residual term, which was not statistically significant (F(8, 184)= 0.40, p= 0.92; η_p^2 = 0.02).

Figure 14. Mean response times (in ms) to words in Experiment 2, by length.

Error rates to words. See Table B6 in Appendix B for means and standard deviations for error rates to words for each Font x Frequency x Length condition. The effect of font homogeneity (t(23)= 0.93, one sided, p= 0.18; η_p^2 = 0.04) was not significant; (for homogeneous fonts, *M* = 0.04. *SD*= 0.06, and for heterogeneous fonts, *M* = 0.05, *SD*= 0.09). For error rates to words, the main effect of word frequency was significant within homogeneous font (t(23)= 2.16, one sided, p= 0.02; η_p^2 = 0.17) and

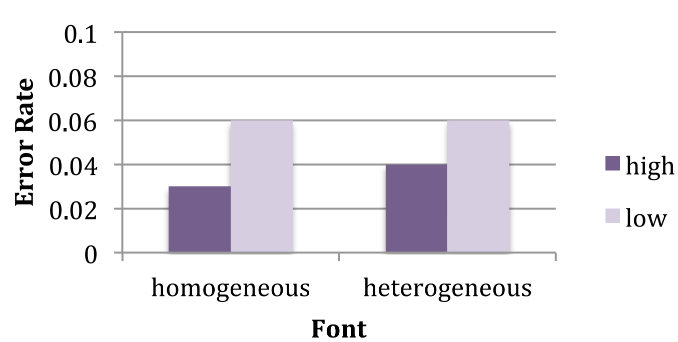

Figure 15. Mean error rates to words in Experiment 2, by frequency.

heterogeneous font (t(23)= 2.00, one sided, p= 0.03; η_p^2 = 0.15). That is, error rate was lower for high frequency words than for low frequency words (for homogeneous, for high frequency: M= 0.03, SD= 0.07; low frequency: M= 0.06, SD= 0.08; for heterogeneous, for high frequency: M= 0.04, SD= 0.08; low frequency: M= 0.06, SD= 0.09) (see Figure 15). The effects of Length within homogeneous font stimuli (t(23)= 0.17, one sided, p= 0.43; η_p^2 = 0.01) and heterogeneous font stimuli (t(23)= 0.04, one sided, p= 0.48; η_p^2 = 0.00007) were not significant. Also, the effects of font within homogeneous font stimuli (F(1, 23)= 0.15, p= 0.70; η_p^2 = 0.006) and heterogeneous font stimuli (F(1, 23)= 0.29, p= 0.59; η_p^2 = 0.01) were not significant. Finally, the residual term was not statistically significant (F(8, 184)= 1.16, p= 0.33; η_p^2 = 0.05).

*Response times to pseudowords***.** See Table B7 in Appendix B for means and standard deviations of response times to pseudowords for each Font x Frequency x Length condition. The effect of font homogeneity was not significant (t(23)= 0.13, one sided, p= 0.45; η_p^2 = 0.0007). For response times to pseudowords, the effect of frequency within homogeneous font (t(23)= 2.65, one sided, p= 0.007; η_p^2 = 0.23) was significant, but the effect of frequency within heterogeneous font was not significant (t(23)= 0.89, one sided, p= 0.19; η_p^2 = 0.03). Results showed that responses to “low frequency” pseudowords were slower than responses to “high frequency” pseudowords (for homogeneous, for “low frequency”: M= 827, SD= 208; “high frequency”: M= 794, SD= 155; for heterogeneous, for “low frequency: M= 816, SD= 173; “high frequency”: M= 805, SD= 173) (see Figure 16). Length was significant within homogeneous font ($t(23)$= 2.98, one sided, p= 0.003; η_p^2 = 0.28) and heterogeneous font ($t(23)$= 5.27, one sided, p= .00001; η_p^2 = 0.55).

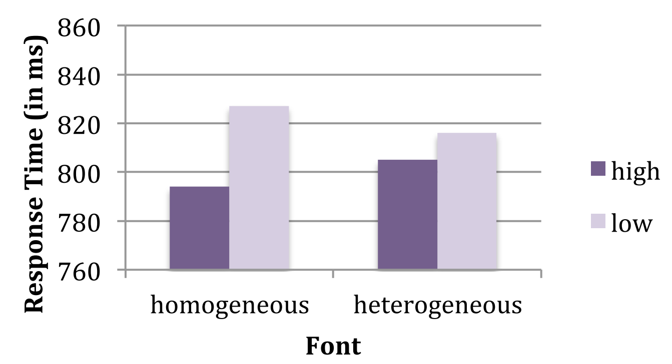

Figure 16. Mean response times (in ms) to pseudowords in Experiment 2, by frequency.

Overall, responses to 6-letter pseudowords were slower than responses to 4-letter pseudowords (6-letter homogeneous: M= 840, SD= 217; 4- letter homogeneous: M= 782, SD= 136; 6-letter, heterogeneous: M= 849, SD= 187; 4- letter, heterogeneous: M= 770, SD= 147 (See Figure 17). The effects of font instantiation within homogeneous font stimuli ($F(1, 23)$= 0.78, p= 0.39; η_p^2 = 0.03) and heterogeneous font stimuli ($F(1, 23)$= 1.11, p= 0.30; η_p^2 = 0.05) were not sigificant. Finally, the residual term was not statistically significant ($F(8, 184)$= 0.44, p= 0.90; η_p^2 = 0.02).

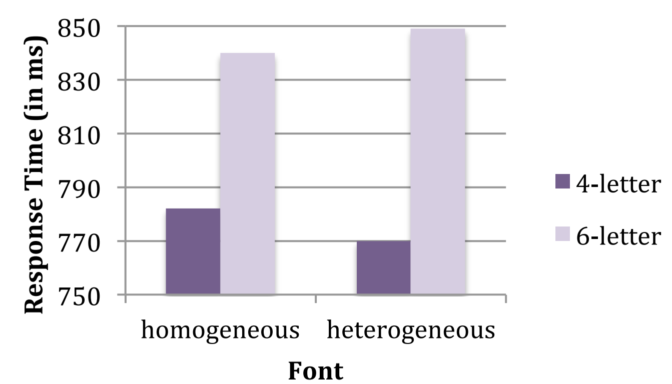

Figure 17. Mean response times (in ms) to pseudowords in Experiment 2, by length.

*Error rates to pseudowords***.** See Table B8 in Appendix B for means and standard deviations for error rates to pseudowords for each Font x Frequency x

Length condition. The effect of font homogeneity was not significant ($t(23)$= 0.63, one sided, p= 0.27; η_p^2 = 0.02). For error rates to pseudowords, frequency within homogeneous font ($t(23)$= 2.61, one sided, p= 0.008; η_p^2 = 0.23) was significant; however, the effect of frequency within heterogeneous font was not significant

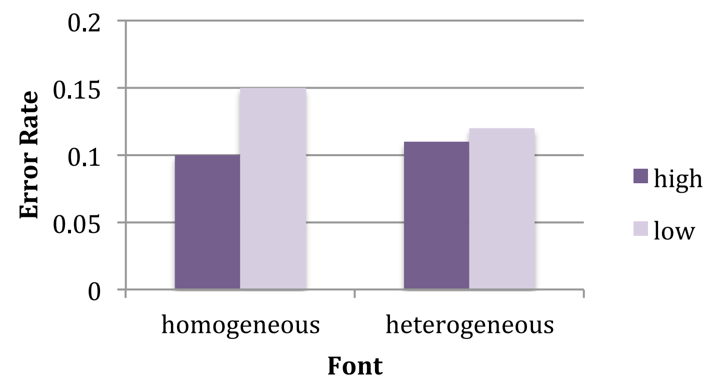

Figure 18. Mean error rates to pseudowords in Experiment 2, by frequency.

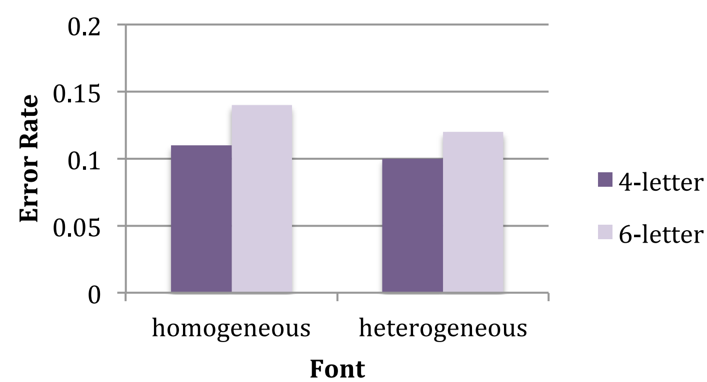

Figure 19. Mean error rates to pseudowords in Experiment 2, by length.

($t(23)$= 0.50, one sided, p= 0.31; η_p^2 = 0.01). For frequency within homogeneous font, error rate was lower on "high frequency" pseudowords than "low frequency" pseudowords (for homogeneous, for "high frequency": M= 0.10, SD= 0.12; "low frequency": M= 0.15, SD= 0.17; for heterogeneous, for "high frequency": M= .11, SD= .13; for "low frequency": M= .12, SD= .14) (See Figure 18). The effect of length within homogeneous font ($t(23)$= 1.72, one sided, p= 0.05; η_p^2 = 0.11) was significant, while the effect of length ($t(23)$= 1.45, one sided, p= 0.08; η_p^2 = 0.08) within heterogeneous font stimuli was not significant. For length within homogeneous fonts, error rate was lower for 4- letter pseudowords than for 6- letter pseudowords (for homogeneous, for 4- letter: *M*= 0.11, *SD*= 0.15; 6- letter: *M*= 0.14, *SD*= 0.15; for heterogeneous, for 4- letter *M*= 0.10, *SD*= 0.12; 6- letter: *M*= 0.12, *SD*= 0.14) (see Figure 19). The effects of font instantiation within homogeneous font ($F(1, 23)= 1.68$, $p= 0.21$; $\eta_p^2 = 0.07$) and heterogeneous font ($F(1, 23)= 1.44$, one sided, $p= 0.24$; $\eta_p^2 = 0.06$) were not significant. Finally, the residual term was not statistically significant ($F(8, 184)= 0.70$, $p= 0.69$; $\eta_p^2 = 0.03$).

Discussion

The significant main effect of font homogeneity for response times to words indicated that a within-item font transition slows lexical decision performance. The significant effect of frequency within both homogeneous font and heterogeneous font showed that, overall, responses to low frequency words were slower than responses to high frequency words. The significant effect of length within heterogeneous font showed that responses to 6- letter items were slower than responses to 4- letter items, and responses were slower, but not significantly so, for 6-letter than 4-letter homogeneous font words. Finally, the comparison of font instantiation within homogeneity and heterogeneity was not significant, which may be attributed to the changes in point-size used for these fonts in this experiment. It may be that any effects of size difference were attenuated by making items in Garamond the same size relative to items in Arial Black. Furthermore, the interactions that were combined into a residual term may show interesting effects that were not examined by the planned contrasts for this thesis; further discussion of these effects is found in the General Discussion below.

CHAPTER IV

GENERAL DISCUSSION

For Experiment 1 the main question was whether responses in the mixed font blocks are slower than responses in pure font blocks. The results of Experiment 1 showed that random trial-to-trial variation in font does not substantially slow lexical decision performance and there were no overall differences between the fonts. Furthermore, the main effects of length and frequency were consistent with previous lexical decision experiments. Although there was no strong evidence for slowed performance, this may be attributed to speed-accuracy trade off. Specifically, for response times to words there was a marginal difference in reaction time in which responses in intermixed font blocks were slightly slower than responses in pure font blocks. However, for error rates to pseudowords there was a significant effect of block that showed lower error rates in intermixed font blocks than in pure font blocks

For Experiment 2 the key question, specifically for words, was whether responses to heterogeneous font items are slower than responses to homogeneous font items. The results of Experiment 2 supported this hypothesis and showed that a within-item font transition slows lexical decision performance. The effects of frequency and length showed results consistent with other lexical decision experiments, in which responses to low frequency items were slower than high frequency items, and responses to 6- letter items were slower than responses to 4-letter items. Although the effect of length was not significant for homogeneous font words in Experiment 2, responses were marginally slower for 6-letter homogeneous font words than for 4-letter ones. In addition, the frequency effect appears bigger for heterogeneous-font than for homogeneous-font words, although given the planned analyses conducted, this cannot be definitively stated. But in general, it seems that for words, the manipulations that slow performance (lower frequency; more letters) had more of an effect for heterogeneous font items than for homogeneous font.

Within homogeneous font and heterogeneous font, no differences were found between the particular fonts used. However, there may be font effects that were not detected in these analyses. For example, in Experiment 2 responses to high frequency words in homogeneous font were slightly slower when presented in Garamond than when presented in Arial Black (Garamond, high frequency: *M*= 656, *SD*= 105; Garamond, low frequency: *M*= 682, *SD*= 112; Arial Black, high frequency: *M*= 652, *SD*= 99; Arial Black, low frequency: *M*= 682, *SD*= 114). Furthermore, for heterogeneous font items, the relative slowness of responses to 6- vs. 4-letter words, and low vs. high frequency words, was larger for items presented in the Arial Black- Garamond transition than for items presented in the Garamond- Arial Black transition (**A**G, 4- letter: *M*= 659, *SD*= 101; **A**G, 6- letter: *M*= 701, *SD*= 165; G**A**, 4-letter: *M*= 655, *SD*= 137; G**A**, 6- letter: *M*= 678, *SD*= 113; **A**G, high frequency: *M*=

654, *SD*= 99; **A**G, low frequency: *M*= 706, *SD*= 165; G**A**, high frequency: *M*= 652, *SD*= 145; G**A**, low frequency: *M*= 700, *SD*= 145). These marginal effects of slowed performance may be attributed to the serifs in Garamond. Although serifs increase the distinctiveness of letters, they are predominantly high spatial frequency information (Arditi & Cho, 2005; Perea, 2013; Perea & Rosa, 2002). Thus, they may require processing by the slower, analytic channels proposed in the multistream model. That is, serifs may serve as visual noise and consequently slow processing when transitioning from a sans serif font (e.g., Arial Black). This is consistent with results from Sanocki (1992), where reaction time was slowed significantly when transitioning from gothic (i.e., sans serif) to serif font. Future research should investigate whether similar results occur with a font transition between two sans serif fonts that differ on several other design characteristics, such as Century Gothic and **Bauhaus** (e.g., do**wn**), or **Impact** and Comic Sans (e.g., **do**wn).

In this thesis, font size was manipulated in order to make items in Garamond the same size relative to items in Arial Black. However, a limitation of this is that the font sizes used (Garamond 36 point, Arial Black 28 point) were larger than those used in other lexical decision experiments. For example, it is possible that a greater effect may have been shown if the items were smaller and the task was more difficult (e.g., Garamond 20 point, Arial Black 16 point) (see Figure 20).

Figure 20. Example of the differences in font size between those used in this thesis (Garamond 36 point, Arial Black 28 point) and a smaller version suggested for future research (Garamond 20 point, Arial Black 16 point).

Although font mixing was associated with poorer performance on a lexical decision task, there may also be positive implications for this effect. For example, one implication lies in the potential use of these effects for the creation of a dysfluent—or completely mixed design—font type for dyslexia. Individuals with dyslexia experience difficulty with reading fluency and comprehension, which leads to difficulty with memory and retrieval of information. It is possible that the use of a dysfluent font would force the reader to process words analytically rather than holistically, which may facilitate reading and comprehension. Furthermore, those who do not have difficulty reading may also benefit from the use of a dysfluent font. That is, if the text is more difficult to read, then it may require more attention and lead to better memory of the information.

Researchers debate whether words are formed on the basis of letter units or on the basis of global word shape. Analytical models suggest that words are formed analytically from component letters, and would predict that performance on heterogeneous font items should not be slower than performance on homogeneous font items. However, others argue that words are processed on the basis of holistic properties. The multistream model suggests that words can be identified from either word-level codes or letter-level codes (Allen et al., 2009). Overall, the results of these experiments provide evidence for the multistream model, which posits that homogeneous font items are processed by the fast, holistic channel and that heterogeneous font items are processed through the slower, analytic channels.

REFERENCES

Allen, P. A., Smith, A. F., Lien, M., Kaut, K. P., & Canfield, A. (2009). A multistream model of visual word recognition. *Attention, Perception, & Psychophysics*, *71*, 281-296. doi:10.3758/APP.71.2.281

Allen, P. A., Smith, A. F., Lien, M, Watt, R. (2015). *Whole Word Shape is in Good Shape.* Manuscript in preparation.

Allen, P. A., Wallace, B., & Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 21, 914-934.

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research*, *45*(23), 2926-2933. doi:10.1016/j.visres.2005.06.013

Besner, D. & Johnston, J.C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.). Lexical representation and process (pp. 291-316). Cambridge, MA: MIT Press.

Day, B., Carter, R., & Meggs, P. (1985). *Typographic Design: Form and Communication*. New York: Van Nostrand Reinhold

Cheng, K. (2005). *Designing type*. New Haven, CT: Yale University Press.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind* 11 (42):220-242.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends In Cognitive Sciences*, *9*(7), 335-341. doi:10.1016/j.tics.2005.05.004

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. W. Walker (Eds.), New approaches to language mechanisms (pp. 257-287). Amsterdam: North-Holland.

Gauthier, I., Wong, A. C., Hayward, W. G., & Cheung, O. S. (2006). Font tuning associated with expertise in letter perception. *Perception*, *35*(4), 541-559. doi:10.1068/p5313

Grainger, J., & Jacobs, A.M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518-565.

Healy, A. F., & Cunningham, T. F. (1992). A developmental evaluation of the role of word shape in word recognition. Memory & Cognition, 20, 141-150.

Humphreys, G. W., Mayall, K., & Cooper, A. G. (2003). The PIG in sPrInG: Evidence on letter grouping from the reading of buried words. *Psychonomic Bulletin & Review*, *10*(4), 939-946. doi:10.3758/BF03196555

Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American Engish.* Providence, RI: Brown Unviersity Press.

McClelland, J.L., & Rumelhart, D.E. (1981). An interaction activation model of context effects in letter perception: Pt. an account of basic findings. *Psychological Review*, 88, 375-407.

Moret-Tatay, C., & Perea, M. (2011). Do serifs provide an advantage in the recognition of written words?. *Journal Of Cognitive Psychology*, *23*(5), 619-624. doi:10.1080/20445911.2011.546781

Perea, M. (2013). Why does the APA recommend the use of serif fonts?. *Psicothema*, *25*(1), 13-17.

Perea, M., & Rosa, E. (2002). Does 'whole-word shape' play a role in visual word recognition?. *Perception & Psychophysics*, *64*(5), 785-794. doi:10.3758/BF03194745

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific schematic tuning. Journal of Experimental Psychology: Human Perception and Performance, 13, 267–278.

Sanocki, T. (1992). Effects of font- and letter-specific experience on the perceptual processing of letters. *The American Journal Of Psychology*, *105*(3), 435-458. doi:10.2307/1423197

Sanocki, T., & Dyson, M. C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention, Perception, & Psychophysics*, *74*(1), 132-145. doi:10.3758/s13414-011-0220-9

Tran, B., Joyce, C.A., & Cottrell, G.W. (2004). Visual expertise depends on how you slice the space. *Proceedings of the 26th Annual Conference of the Cognitive Science Conference.* Mahwah, NJ: Lawrence Erlbaum.

Walker, P. (2008). Font tuning: A review and new experimental evidence. *Visual Cognition*, *16*(8), 1022-1058.

APPENDIX

APPENDIX A

Stimuli Used in Experiments 1 and 2.

Stimulus List for Experiment 1 and Experiment 2

High frequency 4-letter words	High frequency 4-letter pseudowords	Low frequency 4-letter words	Low frequency 4-letter pseudowords	High frequency 6-letter words	High frequency 6-letter pseudowords	Low frequency 6-letter words	Low frequency 6-letter pseudowords
back	dack	root	noot	people	heople	carbon	warbon
much	moch	belt	balt	little	lettle	expert	elpert
well	werl	bare	bame	around	areund	stable	stoble
down	dowy	oral	oras	almost	almest	worker	worrer
each	oach	flew	glew	enough	enoush	lesson	lessin
just	jost	bath	bith	itself	itseld	modest	modesy
good	goud	vein	vean	action	uction	absent	obsent
very	vers	star	stas	either	eather	prayer	proyer
make	dake	pond	tond	across	achoss	warmth	wormth
work	wark	noon	nion	office	offace	define	defone
long	lony	disk	dilk	moment	momert	excuse	excute
here	hert	zero	zera	mother	mothet	mirror	mirroy
both	coth	tail	cail	effect	offect	victim	lictim
life	lofe	slid	shid	except	escept	assist	ansist
same	sace	drug	drog	inside	inride	legend	lenend
used	uset	worn	worp	amount	amoent	cancer	canfer
home	fome	wake	pake	island	islard	saddle	saddye
once	ouce	sink	sonk	indeed	indeer	terror	terrot
high	hiph	dive	dite	simple	timple	arrive	errive
upon	upod	deck	decs	answer	atswer	insure	ilsure
does	soes	bowl	mowl	entire	entore	flower	flawer
away	avay	mess	mest	letter	lether	hungry	huntry
fact	faht	joke	jobe	effort	effost	orange	orante

APPENDIX A
Stimuli Used in Experiments 1 and 2.

High frequency 4-letter words	High frequency 4-letter pseudowords	Low frequency 4-letter words	Low frequency 4-letter pseudowords	High frequency 6-letter words	High frequency 6-letter pseudowords	Low frequency 6-letter words	Low frequency 6-letter pseudowords
hand	hany	gulf	sulf	square	squard	casual	casuar
took	yook	aunt	aont	method	fethod	empire	ompire
head	hoad	urge	urgy	nearly	naarly	refund	rufund
told	tord	ugly	ogly	anyone	ansone	resist	refist
look	looy	rent	rint	nation	natoon	sphere	sphure
knew	gnew	port	pott	record	recold	garage	garafe
city	cety	jail	jain	member	membem	marble	morble
side	sise	diet	niet	normal	cormal	planet	slanet
form	forn	dare	dore	volume	valume	verbal	varbal
felt	nelt	bold	bord	merely	mecely	decent	derent
ever	ener	trap	tral	summer	sumwer	repair	repeir
want	wact	crop	wrop	friend	frield	treaty	treapy
area	areo	boss	biss	former	formen	exceed	exceer
open	apen	slip	slep	chance	shance	string	streng
thus	twus	tray	trag	series	siries	candle	cendle
door	doar	lamp	famp	theory	thiory	humble	hurble
line	lina	lion	liot	myself	mysulf	parlor	parcor
show	thow	kiss	kils	direct	dirent	stolen	stolin
five	feve	chip	chir	spring	sprins	absurd	absurs
gave	bave	taxi	faxi	placed	claced	collar	sollar
seen	seel	soup	saup	degree	dugree	flavor	flevor
week	ceek	lung	fung	manner	mayner	priest	priust
real	rual	flag	flad	larger	larher	stride	strime
love	lote	swim	twim	couple	couphe	parent	parest
full	fult	clue	crue	charge	charke	tennis	tennit
land	tand	ripe	rine	served	lerved	divide	nivide

APPENDIX A

Stimuli Used in Experiments 1 and 2.

High frequency 4-letter words	High frequency 4-letter pseudowords	Low frequency 4-letter words	Low frequency 4-letter pseudowords	High frequency 6-letter words	High frequency 6-letter pseudowords	Low frequency 6-letter words	Low frequency 6-letter pseudowords
feel	fiel	odor	odos	window	wendow	insect	imsect
able	abre	mist	tist	middle	mindle	python	pyshon
type	tupe	fork	fark	appear	appoar	absorb	absurb
soon	hoon	roar	roer	length	lengph	carpet	carpat
mean	mian	grin	gril	fiscal	fiscaw	pepper	peppel
idea	idet	deer	heer	slowly	glowly	purple	gurple
east	eash	acid	alid	corner	cirner	dental	dantal
else	alse	snap	pnap	latter	larter	resort	relort
dead	duad	silk	solk	design	desagn	infant	infont
cold	cond	lens	lecs	choice	choime	prompt	promst
hear	heas	bake	baks	extent	extens	strict	strick
meet	seet	zone	rone	income	ancome	vacant	dacant
hair	huir	calf	celf	expect	enpect	puzzle	pozzle
food	foid	scar	scer	stress	strass	serene	semene
fall	fals	zinc	zint	pretty	prerty	tunnel	tunrel

APPENDIX B

Response Times and Error Rates to

Words and Pseudowords for Experiments 1 and 2.

Table B1. Response time to words
Experiment 1

Block Type	Frequency	Length	Font	*N*	*M*	*SD*
I	H	4	A	24	671	112
			G	24	653	84
		6	A	24	701	114
			G	24	712	140
	L	4	A	24	742	120
			G	24	730	143
		6	A	24	832	358
			G	24	840	230
P	H	4	A	24	677	132
			G	24	668	105
		6	A	24	709	135
			G	24	679	153
	L	4	A	24	725	156
			G	24	702	112
		6	A	24	777	214
			G	24	772	202

Table B2. Error rates to words
Experiment 1

Block Type	Frequency	Length	Font	*N*	*M*	*SD*
I	H	4	A	24	0.01	0.03
			G	24	0	0
		6	A	24	0.04	0.07
			G	24	0.005	0.02
	L	4	A	24	0.07	0.10
			G	24	0.02	0.05
		6	A	24	0.03	0.07
			G	24	0.06	0.09
P	H	4	A	24	0.005	0.02
			G	24	0.01	0.03
		6	A	24	0.01	0.04
			G	24	0.04	0.07
	L	4	A	24	0.05	0.09
			G	24	0.03	0.05
		6	A	24	0.04	0.07
			G	24	0.02	0.06

APPENDIX B
Response Times and Error Rates to
Words and Pseudowords for Experiments 1 and 2.

Table B3. Response time to pseudowords
Experiment 1

Block Type	Frequency	Length	Font	*N*	*M*	*SD*
I	H	4	A	24	952	610
			G	24	914	403
		6	A	24	1011	577
			G	24	1017	396
	L	4	A	24	949	317
			G	24	953	571
		6	A	24	1157	1051
			G	24	1142	992
P	H	4	A	24	924	328
			G	24	919	441
		6	A	24	1036	547
			G	24	974	364
	L	4	A	24	924	272
			G	24	1000	570
		6	A	24	1025	386
			G	24	1058	623

Table B4. Error rates to pseudowords
Experiment 1

Block Type	Frequency	Length	Font	*N*	*M*	*SD*
I	H	4	A	24	0.07	0.11
			G	24	0.05	0.10
		6	A	24	0.02	0.06
			G	24	0.06	0.12
	L	4	A	24	0.06	0.10
			G	24	0.06	0.12
		6	A	24	0.07	0.10
			G	24	0.11	0.12
P	H	4	A	24	0.06	0.10
			G	24	0.10	0.11
		6	A	24	0.07	0.13
			G	24	0.10	0.13
	L	4	A	24	0.09	0.16
			G	24	0.05	0.09
		6	A	24	0.09	0.13
			G	24	0.10	0.09

APPENDIX B

Response Times and Error Rates to

Words and Pseudowords for Experiments 1 and 2.

Table B5. Response time to words
Experiment 2

Font	Frequency	Length	*N*	*M*	*SD*
AA	H	4	24	655	90
		6	24	648	109
	L	4	24	677	106
		6	24	686	124
AG	H	4	24	642	101
		6	24	665	98
	L	4	24	676	101
		6	24	736	208
GA	H	4	24	630	96
		6	24	674	180
	L	4	24	681	165
		6	24	720	120
GG	H	4	24	646	91
		6	24	667	117
	L	4	24	675	116
		6	24	690	111

Table B6. Error rates to words
Experiment 2

Font	Frequency	Length	*N*	*M*	*SD*
AA	H	4	24	0.02	0.08
		6	24	0.02	0.08
	L	4	24	0.07	0.09
		6	24	0.05	0.06
AG	H	4	24	0.03	0.06
		6	24	0.03	0.06
	L	4	24	0.08	0.12
		6	24	0.05	0.09
GA	H	4	24	0.02	0.10
		6	24	0.05	0.07
	L	4	24	0.05	0.08
		6	24	0.04	0.07
GG	H	4	24	0.02	0.04
		6	24	0.04	0.07
	L	4	24	0.05	0.07
		6	24	0.04	0.07

APPENDIX B

Response Times and Error Rates to

Words and Pseudowords for Experiments 1 and 2.

Table B7. Response time to pseudowords
Experiment 2

Font	Frequency	Length	*N*	*M*	*SD*
AA	H	4	24	786	162
		6	24	823	181
	L	4	24	800	115
		6	24	866	310
AG	H	4	24	750	149
		6	24	838	174
	L	4	24	768	139
		6	24	864	194
GA	H	4	24	775	141
		6	24	857	207
	L	4	24	789	164
		6	24	840	178
GG	H	4	24	749	121
		6	24	822	149
	L	4	24	792	143
		6	24	851	213

Table B8. Error rates to pseudowords
Experiment 2

Font	Frequency	Length	*N*	*M*	*SD*
AA	H	4	24	0.08	0.10
		6	24	0.12	0.11
	L	4	24	0.13	0.18
		6	24	0.18	0.17
AG	H	4	24	0.09	0.12
		6	24	0.11	0.13
	L	4	24	0.07	0.12
		6	24	0.11	0.14
GA	H	4	24	0.10	0.11
		6	24	0.12	0.13
	L	4	24	0.13	0.11
		6	24	0.14	0.15
GG	H	4	24	0.06	0.12
		6	24	0.12	0.14
	L	4	24	0.13	0.17
		6	24	0.13	0.15

APPENDIX C
Contrast Tables for Experiments 1 and 2.

Table C1. Contrast Table for Experiment 1.

Variables (and levels): Block Type (Pure vs. Mixed); Font (Arial vs. Garamond); Length (4 vs. 6); Frequency (High vs. Low)

		Pure Font								Mixed Font							
		Arial				Garamond				Arial				Garamond			
		Low		High		Low		High		Low		High		Low		High	
		4	6	4	6	4	6	4	6	4	6	4	6	4	6	4	6
C1	Block Type	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
C2	Font	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1
C3	Block Type x Font	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1
C4	Frequency	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1
C5	Length	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
C6	Freq x Length	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1
C7	Block Type x Frequency	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1
C8	Block Type x Length	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
C9	Font x Freq	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1
C10	Font x Length	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1
C11	Block x Font x Freq	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1
C12	Block x Font x Length	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1
C13	Block x Freq x Length	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1
C14	Font x Freq x Length	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1
C15	Block x Font x Freq x Length	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

APPENDIX C
Contrast Tables for Experiments 1 and 2.

Table C2. Contrast Table for Experiment 2.

Variables (and levels): Stimulus Type (Homogeneous vs. Heterogeneous); Font Instantiation (St) (AA, AG, GA, GG); Length (4 vs. 6); Frequency (High vs. Low)

		Homogeneous								Heterogeneous							
		Arial (AA)				Garamond (GG)				Arial (AG)				Garamond (GA)			
		Low		High		Low		High		Low		High		Low		High	
		4	6	4	6	4	6	4	6	4	6	4	6	4	6	4	6
C1	Homogeneity	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
C2	St within Ho	-1	-1	-1	-1	1	1	1	1	0	0	0	0	0	0	0	0
C3	F within Ho	-1	-1	1	1	-1	-1	1	1	0	0	0	0	0	0	0	0
C4	L within Ho	-1	1	-1	1	-1	1	-1	1	0	0	0	0	0	0	0	0
C5	St x F win Ho	-1	-1	1	1	1	1	-1	-1	0	0	0	0	0	0	0	0
C6	St x L win Ho	-1	1	-1	1	1	-1	1	-1	0	0	0	0	0	0	0	0
C7	F x L win Ho	-1	1	1	-1	-1	1	1	-1	0	0	0	0	0	0	0	0
C8	St x F x L win Ho	-1	1	1	-1	1	-1	-1	1	0	0	0	0	0	0	0	0
C9	St within Het	0	0	0	0	0	0	0	0	-1	-1	-1	-1	1	1	1	1
C10	F within Het	0	0	0	0	0	0	0	0	-1	-1	1	1	-1	-1	1	1
C11	L within Het	0	0	0	0	0	0	0	0	-1	1	-1	1	-1	1	-1	1
C12	St x F win Het	0	0	0	0	0	0	0	0	-1	-1	1	1	1	1	-1	-1
C13	St x L win Het	0	0	0	0	0	0	0	0	-1	1	-1	1	1	-1	1	-1
C14	F x L win Het	0	0	0	0	0	0	0	0	-1	1	1	-1	-1	1	1	-1
C15	St x F x L win Het	0	0	0	0	0	0	0	0	-1	1	1	-1	1	-1	-1	1